

DNA POLYMERASES WITH ENHANCED
LENGTH OF PRIMER EXTENSIONCROSS REFERENCE TO RELATED APPLICATION

5 This application is a continuation-in-part of co-
pending application Serial No. 08/483,535 filed June 7,
1995, which is a continuation-in-part of Patent No.
5,436,149, filed February 19, 1993.

BACKGROUND OF THE INVENTION

The present invention is directed to DNA
polymerases, and more particularly, to a novel
formulation of DNA polymerases, which formulation of
enzymes is capable of efficiently catalyzing the
amplification by PCR (the polymerase chain reaction) of
unusually long and faithful products.

DNA polymerase obtained from the hot springs
bacterium Thermus aquaticus (Taq DNA polymerase) has been
demonstrated to be quite useful in amplification of DNA,
in DNA sequencing, and in related DNA primer extension
techniques because it is thermostable. Thermostable is
defined herein as having the ability to withstand
temperatures up to 95°C for many minutes without becoming
irreversibly denatured, and the ability to polymerize DNA
at high temperatures (60° to 75° C.). The DNA and amino
acid sequences described by Lawyer et al., J. Biol. Chem.
264:6427 (1989), GenBank Accession No. J04639, define the
gene encoding Thermus aquaticus DNA polymerase and the
enzyme Thermus aquaticus DNA polymerase as those terms
are used in this application. The highly similar DNA
polymerase (Tfl DNA polymerase) expressed by the closely
related bacterium Thermus flavus is defined by the DNA
and amino acid sequences described by

SUB
A1

09983002-101501

20

25

30

Akhmetzjanov, A.A., and Vakhitov, V.A. (1992) Nucleic Acids Research 20:5839, GenBank Accession No. X66105. These enzymes are representative of a family of DNA polymerases, also including Thermus thermophilus DNA polymerase, which are thermostable. These enzymes lack a 3'-exonuclease activity such as that which is effective for editing purposes in DNA polymerases such as E. coli DNA polymerase I, and phages T7, T3, and T4 DNA polymerases.

Gelfand et al., U.S. Patent 4,889,818 describe a wild-type (abbreviation used here: WT), native Thermus aquaticus DNA polymerase. Gelfand et al., U.S. Patent 5,079,352 describe a recombinant DNA sequence which encodes a mutein of Thermus aquaticus DNA polymerase from which the N-terminal 289 amino acids of Thermus aquaticus DNA polymerase have been deleted (claim 3 of '352, commercial name Stoffel Fragment, abbreviation used here: ST), and a recombinant DNA sequence which encodes a mutein of Thermus aquaticus DNA polymerase from which the N-terminal 3 amino acids of Thermus aquaticus DNA polymerase have been deleted (claim 4 of '352, trade name AmpliTaq, abbreviation used here: AT). Gelfand et al. report their muteins to be "fully active" in assays for DNA polymerase, but data as to their maximum thermostability is not presented.

Amplification of DNA spans by the polymerase chain reaction (PCR) has become an important and widespread tool of genetic analysis since the introduction of thermostable Taq DNA polymerase for its catalysis. However, one remaining limitation to prior art methods of PCR is the size of the product span that can be amplified. For full-length Taq DNA Polymerase and for N-terminally truncated variants such as Klentaq-278, Klentaq5 and Stoffel Fragment, PCR amplification apparently rapidly becomes inefficient or non-existent as

the length of the target span exceeds 5-6 kb. This was shown even when 30 minutes was used during the extension step of each cycle.

Although there are several reports of inefficient but detectable amplification at 9-10 kb target length and one at 15 kb, most general applications are limited to 5 kb.

Kainze et al. (Analytical Biochem. 202:46-49(1992)) report a PCR amplification of over 10 kb: a 10.9 kb and a 15.6 kb product, utilizing an enzyme of unpublished biological source (commercially available as "Hot Tub" DNA polymerase). Kainze et al. report achieving a barely visible band at 15.6 kb after 30 cycles, starting with 1 ng of λ DNA template per 100 ul of reaction volume. The efficiency of this amplification was shown to be relatively low, although a quantitative calculation of the efficiency was not presented. Attempts by Kainze et al. to make WT Thermus aquaticus DNA polymerase perform in the 10-15 kb size range were not successful, nor have successful results been reported by anyone else for any form of Thermus aquaticus DNA polymerase in this size range.

A DNA polymerase formulation capable of efficient amplification of DNA spans in excess of 6 kb would significantly expand the scope of applications of PCR. For instance, whole plasmids, and constructs the size of whole plasmids, could be prepared with this method, which would be especially valuable in cases in which a portion of the DNA in question is toxic or incompatible with plasmid replication when introduced into E. coli. If this thermostable DNA polymerase preparation simultaneously conferred increased fidelity to the PCR amplification, the resulting large products would be much more accurate, active and/or valuable in research and

applications, especially in situations involving expression of the amplified sequence. If the thermostable DNA polymerase preparation allowed, in addition, more highly concentrated yields of pure product, this would enhance the method of PCR to the point where it could be used more effectively to replace plasmid replication as a means to produce desired DNA fragments in quantity.

SUMMARY OF THE INVENTION

Among the several objects of the invention, therefore, may be noted the provision of a formulation of DNA polymerases capable of efficiently catalyzing primer extension products of greater length than permitted by conventional formulations, including lengths up to at least 35 kilobases, that reduces the mutagenicity generated by the PCR process, particularly in comparison with prior art DNA polymerases and for any target lengths, that maximizes the yield of PCR target fragments and, concomitantly, enhances the intensity and sharpness of PCR product bands, without significant sacrifice in flexibility, specificity, and efficiency; and the provision of an improved process for amplification by PCR which can be utilized to reliably synthesize nucleic acid sequences of greater length and which can effectively utilize PCR products as primers.

Briefly, therefore, the present invention is directed to a novel formulation of thermostable DNA polymerases including at least one thermostable DNA polymerase lacking 3'-5' exonuclease activity and at least one thermostable DNA polymerase exhibiting 3'-5' exonuclease activity.

In another aspect, a kit for the synthesis of a polynucleotide is provided, comprising a first DNA polymerase which possesses 3'-5' exonuclease activity, and a second DNA polymerase which lacks 3'-5' exonuclease activity.

In another aspect, a kit for the synthesis of a polynucleotide is provided, comprising a first DNA polymerase which possesses 3'-5' exonuclease activity, and a second DNA polymerase which lacks 3'-5' exonuclease activity, wherein the first DNA polymerase is selected from the group consisting of Pyrococcus furiosus DNA polymerase, Thermotoga maritima DNA polymerase, Thermococcus litoralis DNA polymerase, and Pyrococcus GB-D DNA polymerase, and the second DNA polymerase

0991002-101501

SUB
A2

is selected from the group consisting of Thermus aquaticus DNA polymerase, (exo-) Thermococcus litoralis DNA polymerase, (exo-) Pyrococcus furiosus DNA polymerase, and (exo-) Pyrococcus GB-D DNA polymerase.

5 In a further embodiment of the invention, a method of amplifying a polynucleotide sequence is provided. The method includes the steps of mixing a composition with a synthesis primer, and a synthesis template, with the composition including a first DNA polymerase possessing 3'-5' exonuclease activity, and a second DNA polymerase lacking 3'-5' exonuclease activity.

10 In yet another aspect of the invention, a method of amplifying a polynucleotide sequence is provided. The method includes the steps of mixing a composition with a synthesis primer, and a synthesis template, with the composition including a first DNA polymerase possessing 3'-5' exonuclease activity which is selected from the group consisting of Pyrococcus furiosus DNA polymerase, Thermotoga maritima DNA polymerase, Thermococcus litoralis DNA polymerase, and Pyrococcus GB-D DNA polymerase, and a second DNA polymerase lacking 3'5 exonuclease activity which is selected from the group consisting of Thermus aquaticus DNA polymerase, (exo-) Thermococcus litoralis DNA polymerase, (exo-) Pyrococcus furiosus DNA polymerase, and (exo-) Pyrococcus GB-D DNA polymerase.

20 Other objects and features will be in part apparent and in part pointed out hereinafter.

SUMMARY OF ABBREVIATIONS AND TERMS

25 The listed abbreviations and terms, as used herein, are defined as follows:

Abbreviations:

30 bp = base pairs
kb = kilobase; 1000 base pairs
nt = nucleotides

BME = beta-mercaptoethanol

PP_i = sodium pyrophosphate

In use, the following 3-letter abbreviations often refer to the single-chain DNA polymerase elaborated by the microorganism.

Pfu = Pyrococcus furiosus

Pwo = Pyrococcus woessii

Taq = Thermus aquaticus

Tfl = Thermus flavus

Tli = Thermococcus litoralis

Klentaq-nnn = N-terminally deleted Thermus aquaticus DNA polymerase that starts with codon nnn+1, although that start codon and the next codon may not match the WT sequence because of alterations to the DNA sequence to produce a convenient restriction site.

WT = wild-type (full length) or deletion of only 3 aa

aa = amino acid(s)

ST = Stoffel fragment, an N-terminal deletion of Thermus aquaticus DNA polymerase that could be named Klentaq-288.

-LA = Long and Accurate; an unbalanced mixture of two DNA polymerases, at least one lacking significant 3'-exonuclease activity and at least one exhibiting significant 3'-exonuclease activity.

PCR = (noun) 1. The Polymerase Chain Reaction

2. One such reaction/amplification experiment.

3. (verb) To amplify via the polymerase chain reaction.

ul = microliter(s)

ATCC = American Type Culture Collection

Megaprimer = double-stranded DNA PCR product used as primer in a subsequent PCR stage of a multi-step procedure.

TKR 2050 PATENT

Deep Vent = DNA polymerase from Pyrococcus species GB-D; purified enzyme is available from New England Biolabs.

Deep Vent exo- = mutant form of Deep Vent DNA polymerase lacking 3'(editing)-exonuclease.

Vent = DNA polymerase from Thermococcus litoralis; purified enzyme is available from New England Biolabs.

Vent exo- = mutant form of Vent DNA polymerase lacking 3'(editing)-exonuclease.

Pfu = DNA polymerase from Pyrococcus furiosus lacking 3'(editing)-exonuclease; purified enzyme is available from Stratagene Cloning Systems, Inc.

Pfu exo- = mutant form of Pfu DNA polymerase ; purified enzyme is available from Stratagene Cloning Systems, Inc.

SEQUENASE = A chemically modified or a mutated form of phage T7 or T3 DNA polymerase wherein the modification or mutation eliminates the 3'-exonuclease activity.

THESIT = polyethylene glycol monododecyl ether.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1 (A-C) are depictions, respectively, of an agarose gel on which was loaded a portion of a test PCR experiment. Figures 1 (A-C) demonstrate the large increase in efficiency of large DNA span PCR achieved by variations of a preferred embodiment of the enzyme formulation of the invention. Although KlenTaq-278 or Pfu DNA polymerase, alone, are shown to catalyze a low level of 6.6 kb PCR product formation, various combinations of the two are seen to be much more efficient. Lower and lower amounts of Pfu in combination with KlenTaq-278 are seen to be effective, down to the minimum presented, 1/640. Of those shown, only a combination of KlenTaq-278 and Pfu can catalyze efficient

amplification of 6.6 kb. Per 100 ul, the indicated level of each enzyme (see Methods, Example 7, for unit concentrations) was used to catalyze PCR reactions templated with 19 ng λ plac5 DNA and primers MBL and MBR. 20 cycles of 94° 2 min., 60° 2 min., 72° 10 min.

Figure 2 is a depiction of an agarose gel on which were analyzed the products of PCR experiments to test the performance of an embodiment of the invention in catalyzing the amplification of fragments even longer than 6.6 kb. Figure 2 demonstrates the ability to amplify 8.4 kb, 12.5 kb, 15 kb, and 18 kb with high efficiency and large yield, utilizing the 1/640 ratio embodiment of the enzyme formulation of the invention. Target product size is indicated above each lane as kb:. Level of template per 100 ul is indicated as ng λ :. 20 or 30 cycles of PCR were each 2 sec. 94°, 11 min. 70°. These early amplifications were non-optimal in several respects compared to the current optimal procedure (see Methods, Example 7): thick-walled tubes were employed instead of thin, catalysis was by 1 ul KlentaqLA-64 (63:1::Klentaq-278:Pfu) instead of KlentaqLA-16, the 27mer primers were used (see Table 3) instead of longer primers, the extension/annealing temperature was 70° instead of 68°, and the Omnigene thermal cycler was used.

Figure 3 is a depiction of an agarose gel of a PCR amplification attempted using a 384 bp megaprimer (double-stranded PCR product) paired with a 43-mer oligonucleotide primer BtV5. Per 100 ul of reaction volume, the following enzymes (see Ex. 7, Methods, for unit concentrations) were used to catalyze amplifications: lane 1, 1 ul Pfu DNA polymerase; lane 2, 1/16 ul Pfu; lane 3, 1 ul Klentaq-278; lane 4, both enzymes together (1 ul Klentaq-278 + 1/16 ul Pfu). The 384 bp band near the bottom of the gel is the megaprimer,

which was originally amplified using Klentaq-278. λ H3 =
lambda DNA digested with HindIII. The only successful
amplification resulted from the combination of the two
enzymes (lane 4). Vent DNA polymerase could substitute
for Pfu with the same result (data not shown).

Figure 4 is a depiction of an agarose gel
demonstrating that 33mers are better than 27mers. Per
100 ul of reaction volume, 2 ng (lanes 1-6) or 10 ng
(lanes 7-12) of lambda transducing phage template were
amplified using 27mer primers (lanes 1-3, 7-9) or 33mer
primers (lanes 4-6, 10-12). Besides being longer, the
33mer lambda primer sequences were situated 100 bp to the
left of primer MBL and 200 bp to the right of primer MBR
on the lambda genome. KlentaqLA-16 in the amounts of
1.2, 1.4, and 1.6 ul was used to catalyze the
amplifications of 12.5, 15, and 18 kb, respectively. 15
ul aliquots (equivalent to 0.3 or 1.5 ng of λ template)
were analyzed by 0.8% agarose electrophoresis.

Figure 5 is a depiction of an agarose gel showing a
CHEF pulse-field analysis (ref. 11, 4 sec. switching
time) of large PCR products amplified by KlentaqLA-16
(1.2 ul) under conditions which were suboptimal with
respect to pH (unmodified PC2 buffer was used) and
thermal cycler (Omnigene). Starting template (see Table
3) was at 0.1 ng/ul and the time at 68° in each cycle was
21 min. for products over 20 kb, 13 min. for lanes 4 & 5,
and 11 min. for lanes 11-14. The volumes of PCR reaction
product loaded were adjusted to result in approximately
equal intensity; in ul: 12,12,4,2; 10,10,10; 2,2,4,1.
The standard size lanes (S) show full-length λ plac5 DNA
(48645 bp) mixed with a HindIII digest of λ DNA. As for
Table 1, the sizes in 5 figures are in base pairs, as
predicted from the primer positions on the sequence of
 λ plac5 DNA, and sizes with decimal points are in kb, as

determined from this gel.

Figure 6 is a depiction of an agarose gel of 28 kb and 35 kb products without (lanes 2,3) and with (lanes 5,6) digestion by restriction enzyme HindIII. Before HindIII digestion, the 28 kb product was amplified with 21 min. extension time per cycle, and the 35 kb product was cycled with 24 min. extension times, both in the RoboCycler at optimum pH (see Ex. 7, Methods). Lanes S (1,4,7) contain markers of undigested λ plac5 and HindIII-digested λ plac5 DNA. Figure 7 is a depiction of an agarose gel showing the results of a Pfu exo^+ mutant test. PCR amplification of 8.4 kb by 30 units (0.7 ug) of Klentaq-278 alone (lanes 1,7) and in combination with a very small admixture (1/16 ul or 1/64 ul, equivalent to 1/6 or 1/25 unit) of archaeobacterial Pfu wild type exo^+ DNA polymerase (+; lanes 2,3) or a mutant thereof lacking the 3'-exonuclease activity (-; lanes 4,5). Lane 6 is the result if 1 ul (2.5 units) of solely Pfu DNA polymerase (wt, exo^+) being employed.

DESCRIPTION OF THE PREFERRED EMBODIMENT

DNA polymerases such as those discussed in this application are commonly composed of up to three identifiable and separable domains of enzymatic activity, in the physical order from N-terminal to C-terminal, of 5'-exonuclease, 3'-exonuclease, DNA polymerase. Taq DNA polymerase has never had a 3'-exonuclease, but certain mutations of its N-terminal portion lead to a deletion of its 5'-exonuclease activity. Other DNA polymerases mentioned, such as Pfu DNA polymerase, do not have the 5'-exonuclease, but their 3'-exonuclease function is central to the aspect of the invention directed to mixtures of DNA polymerases E1 (lacking 3'-exonuclease

activity) and E2 (having 3'-exonuclease activity). In these mixtures, the presence of 5'-exonuclease in either E1 or E2 has not been shown to be essential to the primary advantages of the present invention.

Table 1 below depicts the nucleotide sequence of primers that can be used for amplification of the gene for a preferred embodiment of the DNA polymerase lacking 3'-exonuclease activity (Klentaq-278) included as the primary component of the formulation of DNA polymerases of this invention. The bulk of the DNA sequence for the gene (between the primers) and the resultant amino acid sequence of the enzyme, is defined by the indicated GenBank entry.

Table 1: Primers that amplify the gene for KlenTaq-278

A. The primer at the 5' side of the target fragment. The start codon is indicated with ***.

	NcoI 10	20	30	KT1 36MER									
	GAGCCATGGGCCTCCTCCACGAGTTCGGCCTTCTGG (SEQ ID NO: 1)												

	m	g	L	L	H	E	F	G	L	L	E	...	<-- upper case are WT
aa													
	278	280	282	284	286	288	<-- codon numbering for WT						
aa													
	AGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGG . . . (SEQ ID NO:14)												
	959	969	979	TaqPol.seq GenBank entry									
	Accession No. J04639 (numbering includes 5' non-translated region)												

B. The primer at the 3' side of the target fragment. The two stop codons are indicated with ***. To demonstrate the homology, the other (complementary) strand of the actual primer is shown here.

--other strand-- KLENTAQ32 35mer
HindIII
26 16 ***** 6
GGACTGGCTCTCCGCCAAGGAGTAGTAAGCTTCGC (SEQ ID NO:3)

```

|||||
D W L S A K E *
826 828 830 832
GGACTGGCTCTCCGCCAAGGAGTGATACCACC (SEQ ID NO:15)
2604 2614 2624
TaqPol.seq

```

Table 2 below depicts the nucleotide sequence of the same primers as in Table 1, and shows that these same primers can be used for amplification of the analogous gene from Thermus flavus.

Table 2

The same primers as in Table 1 are homologous to Thermus flavus DNA.

```

10          10          20          30
GAGCCATGGGCCTCCTCCACGAGTTCGGCCTTCTGG      KT1 36MER
*** |||||
m g L L H E F G L L E ... <-- upper case are
WT aa
278 280 282 284 286 288 <-- codon numbering for
WT aa
AGTTTGGGAAGCCTCCTCCACGAGTTCGGCCTCCTGG      Tfl.seq GenBank
entry
1387 1397 1407 Accession number X66105
(numbering includes 5' non-translated region)

26 16 6 --other strand--
GGACTGGCTCTCCGCCAAGGAGTAGTAAGCTTCGC      KLENTAQ32 35mer
|||||
D W L S A K E *
826 828 830
GGACTGGCTCTCCGCCAAGGAGTAGGGGGGTCCTG      Tfl.seq
3032 3042 3052

```

Referring now to Table 1, the primers and logic for amplification by PCR of the recombinant DNA sequence encoding a preferred embodiment of the thermostable DNA polymerase of the invention lacking 3'-exonuclease activity (referred to herein as Klentaq-278), are set forth. As depicted in Table 1, an initiator methionine

5 and a glycine residue occupy the first two N-terminal positions of Klentaq-278, previously occupied by residues 279 and 280 of WT Thermus aquaticus DNA polymerase, followed by the amino acid sequence of wild-type Thermus aquaticus DNA polymerase, beginning with the amino acid residue at position 281 as described by Lawyer et al. The codons encoding amino acid residues 1 through 280 of Thermus aquaticus DNA polymerase are therefore deleted, and the amino acids 1 thru 280 are not present in the resulting gene product.

The primers and logic for amplification of another preferred embodiment of the DNA polymerase of the invention lacking 3'-exonuclease activity are set forth in Table 2. In this embodiment, the same deletion mutation described above is made to the highly analogous enzyme Thermus flavus DNA polymerase.

20 The mutant DNA polymerase Klentaq-278 exhibits thermostability at temperatures above those reported for previous variants of Thermus aquaticus DNA polymerase and has demonstrated a fidelity in final PCR products which is greater than that of WT Thermus aquaticus DNA polymerase, when both are utilized at the 72° C temperatures recommended for DNA synthesis.

25 A vector is also provided which includes a recombinant DNA sequence encoding a DNA polymerase comprising the amino acid sequence of Thermus aquaticus or Thermus flavus DNA polymerase, except that it adds a methionine and glycine residue at the N-terminal and excludes the N-terminal 280 amino acids of wild-type Thermus aquaticus DNA polymerase (see Lawyer et al., supra).

30 In preferred embodiments, the vector is that nucleic acid present as plasmid pWB254b (SEQ ID NO:5) deposited as ATCC No. 69244 or a host cell containing such a

09981002 "301701

vector.

5 In a related aspect, the invention features purified DNA polymerases of the type described herein. As used in this application, "purified" means that the polymerase of the invention is isolated from a majority of host cell proteins normally associated with it. Preferably, the polymerase is at least 10% (w/w) of the protein of a preparation. Even more preferably, it is provided as a homogeneous preparation, e.g., a homogeneous solution.

10 In general, the recombinant DNA sequence encoding for a preferred DNA polymerase lacking 3'-exonuclease activity which serves as the primary component of the DNA polymerase formulation of the present invention is amplified from a Thermus aquaticus genomic DNA or from a clone of the portion of the Thermus aquaticus DNA polymerase gene which is larger than the desired span, using the polymerase chain reaction (PCR, Saiki et al., Science 239:487, 1988), employing primers such as those in Table 1 into which appropriate restriction sites have
20 been incorporated for subsequent digestion.

25 The recombinant DNA sequence described above is then cloned into an expression vector using procedures well known to those in this art. Specific nucleotide sequences in the vector are cleaved by site-specific restriction enzymes such as NcoI and HindIII. Then, after optional alkaline phosphatase treatment of the vector, the vector and target fragment are ligated together with the resulting insertion of the target codons in place adjacent to desired control and
30 expression sequences. The particular vector employed will depend in part on the type of host cell chosen for use in gene expression. Typically, a host-compatible plasmid will be used containing genes for markers such as ampicillin or tetracycline resistance, and also

containing suitable promoter and terminator sequences.

In a preferred procedure, the recombinant DNA expression sequence described above is cloned into plasmid pWB253 (expresses KlenTaq-235 deposited as ATCC No. 68431) or pWB250 (expresses luciferase/NPTII fusion), the backbone of which is pTAC2 (J.Majors, Washington University), a pBR322 derivative. The specific sequence of the resulting plasmid, designated pWB254b is SEQ ID NO: 5.

Bacteria, e.g., various strains of E. coli, and yeast, e.g., Baker's yeast, are most frequently used as host cells for expression of DNA polymerase, although techniques for using more complex cells are known. See, e.g., procedures for using plant cells described by Depicker, A., et al., J. Mol. Appl. Gen. (1982) 1:561. E. coli host strain X7029, wild-type F⁻, having deletion X74 covering the lac operon is utilized in a preferred embodiment of the present invention.

A host cell is transformed using a protocol designed specifically for the particular host cell. For E. coli, a calcium treatment, Cohen, S.N., Proc. Natl. Acad. Sci. 69:2110 (1972), produces the transformation. Alternatively and more efficiently, electroporation of salt-free E. coli is performed after the method of Dower et al. (1988), Nucleic Acids Research 16:6127-6145. After transformation, the transformed hosts are selected from other bacteria based on characteristics acquired from the expression vector, such as ampicillin resistance, and then the transformed colonies of bacteria are further screened for the ability to give rise to high levels of isopropylthiogalactoside (IPTG)-induced thermostable DNA polymerase activity. Colonies of transformed E. coli are then grown in large quantity and expression of KlenTaq-278 DNA polymerase is induced for

isolation and purification.

Although a variety of purification techniques are known, all involve the steps of disruption of the E. coli cells, inactivation and removal of native proteins and precipitation of nucleic acids. The DNA polymerase is separated by taking advantage of such characteristics as its weight (centrifugation), size (dialysis, gel-filtration chromatography), or charge (ion-exchange chromatography). Generally, combinations of these techniques are employed together in the purification process. In a preferred process for purifying Klentaq-278 the E. coli cells are weakened using lysozyme and the cells are lysed and nearly all native proteins are denatured by heating the cell suspension rapidly to 80° C and incubating at 80-81° C for 20 minutes. The suspension is then cooled and centrifuged to precipitate the denatured proteins. The supernatant (containing Klentaq-278) then undergoes a high-salt polyethylene-imine treatment to precipitate nucleic acids. Centrifugation of the extract removes the nucleic acids. Chromatography, preferably on a heparin-agarose column, results in nearly pure enzyme. More detail of the isolation is set forth below in Example 3.

The novel DNA polymerase formulation of the present invention may be used in any process for which such an enzyme formulation may be advantageously employed. This enzyme formulation is particularly useful for PCR amplification techniques, but may also be used for other processes employing DNA polymerases such as nucleic acid sequencing, cycle sequencing, DNA restriction digest labelling and blunting, DNA labelling, in vivo DNA footprinting, and primer-directed mutagenesis.

Amplification

Polymerase chain reaction (PCR) is a method for

5 rapidly amplifying specific segments of DNA, in geometric progression, up to a million fold or more. See, e.g., Mullis U.S. Patent No. 4,683,202, which is incorporated herein by reference. The technique relies on repeated cycles of DNA polymerase-catalyzed extension from a pair of primers with homology to the 5' end and to the complement of the 3' end of the DNA segment to be amplified. A key step in the process is the heat denaturing of the DNA primer extension products from their templates to permit another round of amplification. The operable temperature range for the denaturing step generally ranges from about 93°C to about 95°C, which irreversibly denatures most DNA polymerases, necessitating the addition of more polymerase after each denaturation cycle. However, no additional DNA polymerase needs to be added if thermostable DNA polymerases such as Thermus aquaticus DNA polymerase are used, since they are able to retain their activity at temperatures which denature double-stranded nucleic acids. As described in Example 4, below, KlenTaq-278 has demonstrated the ability to survive meaningful repeated exposure to temperatures of 99°C, higher than for any previously known DNA polymerase.

25 Deposit

30 Strain pWB254b/X7029 was deposited with the American Type Culture Collection, Maryland, on February 18, 1993 and assigned the number ATCC 69244. Applicant acknowledges his responsibility to replace this culture should it die before the end of the term of a patent issued hereon, 5 years after the last request for a culture, or 30 years, whichever is the longer, and his responsibility to notify the depository of the issuance

of such a patent, at which time the deposits will be made available to the public. Until that time the deposits will be made available to the Commissioner of Patents under the terms of 37 C.F.R. Section 1-14 nad 35 U.S.C. §112.

In the principal aspect of the invention, a target length limitation to PCR amplification of DNA has been identified and addressed. Concomitantly, the base pair fidelity, the ability to use PCR products as primers, and the maximum yield of target fragment were increased. These improvements were achieved by the combination of a DNA polymerase lacking significant 3'-exonuclease activity, preferably, KlenTaq-278 described above, with a low level of a DNA polymerase exhibiting significant 3'-exonuclease activity (for example, Pfu, Vent, or Deep Vent). Surprisingly, target fragments of at least 35 kb can be amplified to high yields from, for example, 1 ng lambda DNA template with this system.

Moreover, products in the range 6.6 to 8.4 kb can be efficiently amplified by a formulation of thermostable DNA polymerases consisting of a majority component comprised of at least one thermostable DNA polymerase lacking 3'-exonuclease activity and a minority component comprised of at least one thermostable DNA polymerase exhibiting 3'-exonuclease activity, i.e., wherein the ratio of DNA polymerase lacking 3'-exonuclease activity to that exhibiting 3'-exonuclease activity exceeds 1 to 1, measured by DNA polymerase activity units (or by weight where the DNA polymerase activity of the 3'-exonuclease activity-exhibiting enzyme has been eliminated, as described below).

The prior art technology only allowed relatively inefficient and sporadic amplification of fragments in this size range, resulting in only relatively faint

5 product bands or no detectable product at all. In light of the current discovery, I believe I understand the reason for the inefficiency of the prior art. Without limiting myself to any particular theory, it is believed that Thermus aquaticus DNA polymerase and its variants are slow to extend a mismatched base pair (which they cannot remove since they lack any 3'-exonuclease). A couple of companies (New England Biolabs and Stratagene) have introduced thermostable enzymes which exhibit a 3'-(editing) exonuclease which should, one would think, allow the removal of mismatched bases to result in both efficient extension and more accurately copied products. In practice, these two enzymes (Vent and Pfu DNA polymerase) are unreliable and much less efficient than expected. One possible explanation for the unreliability of these enzymes for PCR is that the 3'-exonuclease often apparently attacks and partially degrades the primers so that little or no PCR is possible. This primer attack problem is worse for some primers than others. It has been reported (Anonymous, The NEB Transcript, New England Biolabs, (March, 1991) p. 4.) that the Vent DNA polymerase leaves the 5' 15 nt intact, so that if the annealing conditions allow that 15 nt to prime, PCR could presumably proceed. This would of course only allow annealling at lower, non-selective temperatures, and the 5' 15 nt of the primers must be exactly homologous to the template.

20 I have discovered that the beneficial effects of a 3'-exonuclease can be obtained with an unexpectedly minute presence of one or more DNA polymerases which exhibit 3'-exonuclease activity (herein called "E2") such as certain Archaeobacterial DNA polymerases, whilst efficient extension is being catalyzed by a large amount of one or more DNA polymerases which lack 3'-exonuclease

25

30

activity, such as KlenTaq-278 or AT (herein called "E1"). As a minority component of a formulation or mixture of DNA polymerases, the unreliability and inefficiency of the 3'-exonuclease DNA polymerase, discussed above, is substantially reduced or eliminated. Moreover, since it is believed that the 3'-exonuclease is removing mismatches to eliminate pausing at the mismatches, the resulting DNA exhibits fewer base pair changes, which is a valuable decrease in the mutagenicity of PCR without sacrificing flexibility, specificity, and efficiency. In fact, the combination, even for KlenTaq-278/Pfu units ratios as high as 2000, exhibited greatly increased efficiency of amplification. For most applications, the mixture of DNA polymerases must be at a relative DNA polymerase unit ratio of E1 to E2 of at least about 4:1, before enhanced product length and yield can be achieved. When Pfu DNA polymerase was used in the formulation, the ratio preferably is in the range 80 to 1000 parts KlenTaq-278 per part (unit) Pfu, more preferably from about 150 to about 170:1, and most preferably, is about 160:1, depending somewhat on the primer-template combination. Similar ratios are preferred for mixtures of Pfu and KlenTaq-291.

If Deep Vent is substituted for Pfu for use in combination with KlenTaq-278 or -291, the most preferred ratios for most applications increases to from about 450 to about 500:1 E1 to E2; if full-length (WT) Taq or Amplitaq is included as E1, the most preferred ratio to Pfu or other E2 component is between about 10 and about 15:1 of E1 to E2.

E2 of the invention includes, but is not limited to, DNA polymerase encoded by genes from Pfu, Vent, Deep Vent, T7 coliphage, Tma, or a combination thereof. E1 of the invention includes, but is not limited to, a mutant,

3'-exonuclease negative form of an E2 DNA polymerase, or alternatively, a DNA polymerase which, in unmutated form, does not exhibit significant 3'-exonuclease activity, such as the DNA polymerases encoded by genes from Taq, Tfl, or Tth, or a combination thereof.

As discussed below, the formulation of DNA polymerases of the present invention also includes formulations of DNA polymerase wherein E1 comprises a reverse transcriptase such as SEQUENASE.

Additional examples of the formulations of the present invention include mixtures wherein E1 comprises or consists of a mutant or chemical modification of T7 or T3 DNA polymerase and E2 comprises or consists of a wild-type T7 or T3 DNA polymerase, or, in another variation, E1 comprises or consists of a Vent DNA polymerase lacking 3'-exonuclease activity (sold by New England Biolabs as Vent exo-) and E2 comprises or consists of Vent.

The principal here discovered, namely the use of low levels of 3' exonuclease during primer extension by a DNA polymerase lacking 3' exonuclease, is preferably employed using thermostable DNA polymerases, but is applicable to general DNA polymerase primer extensions, including normal temperature incubations (i.e. using non-thermostable DNA polymerases) and including reverse transcriptase enzymes, which are known to lack a 3'-(editing) exonuclease (Battula & Loeb, 1976). An example of the former is the use of SEQUENASE (exo-) as the majority enzyme, and wild-type T7 DNA polymerase (exo+) or Klenow fragment as the minority component. An example of the latter is AMV (Avian Myoblastosis Virus) or MLV (Murine Leukemia Virus) Reverse Transcriptase as the major component, and Klenow fragment, T7 DNA polymerase, or a thermostable DNA polymerase such as Pfu or Deep Vent as the minor component. Because of the lower activity of

thermostable DNA polymerases at the temperatures of 37 degrees and 42 degrees used by these reverse transcriptases, higher levels are likely to be required than are used in PCR. Although Klenow fragment DNA polymerase is not a preferred DNA polymerase using RNA as a template, it does function to recognize this template (Karkas, 1973; Gulati, Kacian & Spiegelman, 1974), particularly in the presence of added Mn ion. Added Mn ion is routinely used to achieve reverse transcription by thermostable DNA polymerase Tth, unfortunately (in the prior art) without the benefit of an exo+ component. It must be stressed that for the use of the exo+ component for reverse transcriptase reactions, extra care must be taken to ensure that the exo+ component is entirely free of contaminating RNase.

The following references describe methods known in the art for using reverse transcriptases, and are hereby incorporated by reference.

Battula N. Loeb LA. On the fidelity of DNA replication. Lack of exodeoxyribonuclease activity and error-correcting function in avian myeloblastosis virus DNA polymerase. Journal of Biological Chemistry. 251(4):982-6, 1976 Feb 25.

Gulati SC. Kacian DL. Spiegelman S. Conditions for using DNA polymerase I as an RNA-dependent DNA polymerase. Proceedings of the National Academy of Sciences of the United States of America. 71(4):1035-9, 1974 Apr.

Karkas JD. Reverse transcription by Escherichia coli DNA polymerase I. Proc Natl Acad Sci U S A. 70(12):3834-8, 1973 Dec.

**DNA Polymerase with no polymerase activity, only 3'-
exonuclease activity:**

While not limiting myself to a particular theory, applicant believes that the enzymatic activity of value in the minor (E2) component is the 3'-exonuclease activity, not the DNA polymerase activity. In fact, it is further believed that this DNA polymerase activity is potentially troublesome, leading to unwanted synthesis or less accurate synthesis under conditions optimized for the majority (E1) DNA polymerase component, not the minority one. As taught by [Bernad, Blanco and Salas (1990) Site-directed mutagenesis of the YCDTDS amino acid motif of the phi 29 DNA polymerase, Gene 94:45-51.] who mutated the "Region I" DNA conserved DNA polymerase motif of phi 29 DNA polymerase, either Region III or Region I of the Pfu DNA polymerase gene are mutated, which has been sequenced by Uemori, T., Ishino, Y., Toh, H., Asada, F. and Kato, I. Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*, Nucleic Acids Res. 21, 259-265 (1993).

The following examples illustrate the invention.

EXAMPLE 1

Construction of an Expressible Gene for Klentaq-278

In order to construct the Klentaq-278 DNA polymerase gene having a recombinant DNA sequence as described above, the following procedure was followed.

The mutated gene was amplified from 0.25 ug of total *Thermus aquaticus* DNA using the polymerase chain reaction (PCR, Saiki et al., Science 239:487, 1988) primed by the two synthetic DNA primers of Table 1. Primer KT1, SEQ ID NO:1, has homology to the wild-type DNA starting at codon 280; this primer is designed to incorporate a NcoI site into the product amplified DNA. Primer Klentaq32, SEQ ID

NO:3, a 33mer spanning the stop codon on the other strand of the wild-type gene encoding Thermus aquaticus DNA polymerase, and incorporating a HindIII site and a double stop codon into the product DNA.

The buffer for the PCR reaction was 20 mM Tris HCl pH 8.55, 2.5 mM MgCl₂, 16 mM (NH₄)₂SO₄, 150 ug/ml BSA, and 200 uM each dNTP. The cycle parameters were 2' 95°, 2' 65°, 5' 72°.

In order to minimize the mutations introduced by PCR (Saiki et al., supra), only 16 cycles of PCR were performed before phenol extraction, ethanol precipitation, and digestion with the restriction enzymes NcoI and HindIII.

EXAMPLE 2

Preparation of an Expression Vector

The product NcoI and HindIII fragment was cloned into plasmid pWB254b which had been digested with NcoI, HindIII, and calf intestine alkaline phosphatase. The backbone of this plasmid, previously designated pTAC2 and obtained from J. Majors, carries the following elements in counter-clockwise direction from the PvuII site of pBR322 (an apostrophe ' designates that the direction of expression is clockwise instead of counter clockwise): a partial lacZ' sequence, lacI', lacPUV5 (orientation not known), two copies of the tac promoter from PL Biochemicals Pharmacia-LKB; catalog no. 27-4883), the T7 gene 10 promoter and start codon modified to consist of a NcoI site, a HindIII site, the trpA terminator (PL no. 27-4884-01), an M13 origin of replication, and the Amp^R gene of pBR322. Expression of the cloned gene is expected to be induced by 0.1 mM IPTG.

Ampicillin-resistant colonies arising from the cloning were assayed by the single colony thermostable DNA

polymerase assay of Sagner et al. (1991) [GENE 97:119-23] and 4 strong positives were sized by the toothpick assay (Barnes, Science 195:393, 1977). One of these, number 254.7, was of the expected size except for a small proportion of double insert. This plasmid was further purified by electroporation into E. coli X7029 and screened for size by the toothpick assay, and one plasmid of the expected size with no double insert contamination was designated pWB254b. This plasmid was used for the production of Klentaq-278 described herein.

EXAMPLE 3

Purification of Large Amounts of Klentaq-278

Plasmid pWB254 has a double (tandem repeat) tac promoter and the T7 gene 10 leader sequence, an ATG start codon, a glycine codon and then codons 280-832 of Thermus aquaticus DNA polymerase, then a tandem pair of stop codons followed by the trp transcription terminator. The pBR322-based plasmid vector (pTac2 from John Majors) is ampicillin resistant. The cells are grown on very rich medium (see below). Bacterial host X7029 is wild-type F⁻ E. coli except for deletion X74 of the lac operon.

Medium: Per liter water, 100 mg ticarcillin (added when cool), 10 g Y.E., 25 g. Tryptone, 10 g. glucose, 1XM9 salts with no NaCl (42 mM Na₂PO₄, 22 mM KH₂PO₄, 19 mM NH₄Cl). Do not autoclave the glucose and the 10XM9 together; instead, autoclave one of them separately and mix in later. Adjust pH to 8 with 5 M NaOH (about 1 ml). Add IPTG to 0.1 mM at OD₅₅₀ = 1 or 2, and shake well at 30° C. From OD = 2 up to 8 or 10, every half hour or so do the following:

1. Read the pH with pH sticks 5-10. Adjust to pH 8.5 with 5 M NaOH and swirling (2 to 5 ml per liter) whenever the pH falls below 8.

2. Read and record the OD_{550} , usually as a 1/10 or 1/50 dilution.

3. This addition of glucose is optional and not necessarily of any value (evaluation of this question is incomplete at this time.) Read the glucose level with glucose sticks, and add an additional 0.5% (10 ml of 50%) if the level falls below 0.2%.

If it is late, the cells can shake at 30° C all night after the last pH adjustment. Alternatively, set them in the cold room if they have not grown much in a few hours.

Concentrate the cells e.g. by centrifugation in a GS3 rotor for 8 minutes at 8 krpm. Pour off the supernatant and add culture to spin more down onto the same pellets.

Lysis:

Resuspend the cells in milliliters of TMN buffer equal to twice the packed cell weight in grams: (50 mM Tris-HCl pH 8.55, 10 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$).

To each 300 ml of cell suspension add 60 mg lysozyme and incubate the cells at 5-10° C. with occasional swirling for 15 minutes. Then add NP40 or Triton X100 to 0.1%, and Tween 20 to 0.1%, by adding 1/100 volume of a solution of 10% in each. Then heat the cell suspension rapidly to 80° C. by swirling it in a boiling water bath, then maintain the cells (fast becoming an extract) at 80-81° C. for 20 minutes. Use a clean thermometer in the cells to measure temperature. Be sure the flask and bath are covered, so that even the lip of the flask gets the full heat treatment. After this treatment, which is expected to have inactivated all but a handful of enzymes, cool the extract to 37° C. or lower in an ice bath and add 2 ml of protease inhibitor (100 mM PMSF in isopropanol). From this point forward, try not to con-

5 tact the preparation with any flask, stir bar, or other object or solution that has not been autoclaved. (Detergents and BME are not autoclavable. The PEI and ammonium sulfate are also not autoclaved.) The purpose of the autoclaving is not only to avoid microbial contamination, but also to avoid contamination with DNA or nucleases.

Distribute into centrifuge bottles and centrifuge at 2° C. (for instance, 30 minutes at 15 krpm in a Sorval SS-34 rotor or 14 h at 4 krpm in a GS3 rotor). The supernatant is designated fraction I, and can be assayed for DNA polymerase activity.

High-salt PEI precipitation

0581002
15
20 After rendering fraction I 0.25 M in NaCl (add 14.6 g per liter), add five percent Polymin-P (PEI, polyethylene-imine, Sigma) dropwise with stirring on ice to precipitate nucleic acids. To determine that adequate Polymin-P has been added, and to avoid addition of more than the minimum amount necessary, test 1/2 ml of centrifuged extract by adding a drop of Polymin-P, and only if more precipitate forms, add more Polymin-P to the bulk extract, mix and retest. Put the test aliquots of extract back into the bulk without contaminating it.

25 To confirm that enough PEI has been added, centrifuge 3 ml and aliquot the supernatant into 1/2 ml aliquots. Add 0, 2, 4, 6 or 10 ul of 5% PEI. Shake, let sit on ice, and centrifuge in the cold. Load 15 ul of these aliquot supernatants onto an agarose gel containing ethidium bromide and electrophorese until the blue dye has travelled 2 cm. Inspect the gel on a UV light box
30 for detectable DNA or RNA in the supernatant. For the bulk extract, use about 1/100 volume (i.e. 2-3 ml for a 300 ml extract) excess 5% PEI over the minimum necessary to remove all DNA by the agarose gel test.

Stir in the cold for at least 15 minutes.

Centrifugation of the extract then removes most of the nucleic acids. Keep the supernatant, avoiding any trace of the pellet.

Dilute the PEI supernatant with KTA buffer until the conductivity is reduced to at or below the conductivity of KTA buffer with added 22 mM ammonium sulfate. (Check conductivity of 1/40 dilution compared to similar dilution of genuine 22 mM A.S. in KTA.) Usually this is about a 5-fold dilution.

Chromatography with Bio-Rex 70 (used by Joyce & Grindley) (Joyce, C.M. & Grindley, N.D.E. (1983) Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA polymerase I of E. coli, Proc. Natl. Acad. Sci. U.S.A. 80, 1830-1834) is unsuccessful (no binding), but unavoidable, since without it, the next column (heparin agarose) will not work efficiently. I believe that the important function of the Bio-Rex 70 step is to remove all excess PEI, although it is possible that some protein is removed as well. CM-cellulose does not substitute for Bio-Rex 70.

Pass the diluted PEI supernatant through equilibrated Bio-Rex 70 (10 ml per 100 g. cells). The polymerase activity flows through. Rinse the column with 2 column volumes of 22 mM A.S. / KTA. Our procedure is to set up the following heparin agarose column so that the effluent from the Bio-REX 70 column flows directly onto it.

Heparin Agarose Chromatography (room temperature, but put fractions on ice as they come off.)

Load the Bio-Rex flow-through slowly onto heparin agarose (Sigma; 10 ml per 100 grams of cells [this could be too little heparin agarose].) Wash with several column volumes of KTA + 22 mM A.S., then three column volumes of KTA + 63% glycerol + 11 mM A.S., then elute the

pure enzyme with KTA + 63% glycerol + 222 mM A.S. + 0.5% THESIT (this is more THESIT for the final eluate.)

Pool the peak of polymerase activity or OD₂₈₀/ (starts about at 2/3 of one column volume after 222 mM starts, and is about 2 column volumes wide). Store pool at -20° C.

The storage buffer is a hybrid of, and a slight variation of, AmpliTaq storage buffer as recommended by Perkin-Elmer Cetus and Taq storage buffer used by Boehringer-Mannheim: 50% glycerol (v/v; 63% w/v), 222 mM ammonium sulfate (diluted to about 50 mM for bench-strength samples), 20 mM Tris-HCl pH 8.55, 0.1 mM EDTA, 10 mM mercaptoethanol, 0.5% THESIT).

The THESIT causes some thickening and cloudiness below -10° C. This seems to cause no harm, but we suggest you warm the enzyme to 0° C. on ice before aliquoting for use. THESIT replaces the combination of 0.5% Triton-X100, 0.5% Tween 20, which you may want to consider as an alternative.

I have had sporadic reports that freezing can inactivate the enzyme. Exercise caution in this regard. This question is under current investigation. Storage at -80° (after quick-cooling with liquid nitrogen) is being tested and looks promising, but more than one freeze-thaw cycle has been deleterious to the enzyme preparation on some occasions.

Our final yield of enzyme from 7 liters (100 g cells) was once 28 ml at a concentration of 120,000 units per ml (4 x bench-strength).

1/4 ul of bench-strength enzyme will support the PCR of a 2 kb span of DNA in a 100 ul reaction. Template is 5-10 ng of plasmid DNA. Each cycle consists of 1 min 98° C, 1 min 65° C, 6 min 72° C.

Cycle number is 16-20. Less enzyme is needed for smaller-sized products (1/8 ul for 500 bp) and more enzyme is needed for larger products (1 ul for 5 kb).

5

KTA Buffer**per liter**-----
20 mM Tris 8.55-----
10 ml of 2 M

10 mM BME

0.7 ml neat

10% w/v Glycerol

100 g.

0.1 mM EDTA

0.2 ml of .5 M

0.1% w/v THESIT

10 ml of 10%

Rough Incorporation Assay

1 X PC2 Buffer (20 mM Tris-HCl pH 8.55, 2.5 mM MgCl₂,
16 mM (NH₄)₂SO₄, 100 ug/ml BSA)

200-250 ug/ml activated salmon sperm DNA

40 uM each dNTP + 10-50 uCi α-³²P-dATP per ml

To 25 ul assay mix on ice add 0.2 ul of enzyme fraction, undiluted, or diluted in 8 ul of 1XPC2 buffer (or a 1/5 or 1/25 dilution thereof.)

Prepare standard Klentaq or Amplitaq, zero enzyme and total input samples, also. Incubate 10 min. at 72° C., then chill. Spot 5 or 8 ul onto filter paper and wash twice for 5 - 10 min. with 5% TCA, 1% PP_i. If pieces of paper were used, count each using Cerenkov radiation or hand monitor. If a single piece of 3 MM paper was used, autoradiograph for 60'.

PCR Assay to give 2 kb product.

Make up 1 ml of PCR reaction containing 50 ng of plasmid pLc (a clone of an R color control cDNA from maize. PNAS 86:7092; Science 247:449), 200

FD 1002-10101

20

25

30

pmoles each of primers Lc5 (SEQ ID NO:11) and Lc3 (SEQ ID NO:12), PC2 buffer and 200 μ M dNTPs, but no enzyme.

Distribute 100 μ l into tube one, and 50 μ l into the rest of 8-10 tubes. Add 1 μ l of final pool of KlenTaq to tube one and mix. Then remove 50 μ l to tube two and mix that, and so on down the series, which will then contain decreasing amounts of enzyme in two-fold steps. Cover each 50 μ l reaction with a drop of mineral oil, spin, and PCR 16 cycles at 2' 95° C, 2' 65° C, 5' 72° C.

Final Bench-Strength KlenTaq-278 Enzyme

Using 63% glycerol / KTA (.5% THESIT) buffer with 222 mM ammonium sulfate, dilute the pool conservatively so that 1/4 μ l should easily catalyze the amplification the 2 kb span by PCR. Do not decrease the ammonium sulfate concentration below 50 mM. Store at -20° C.

EXAMPLE 4

A PCR amplification assay to produce 2 kb of DNA product was conducted using Thermus aquaticus DNA polymerase (AmpliTaq) and KlenTaq-278. To test polymerase thermostability at elevated temperatures, the DNA denaturation step of the PCR amplification reactions were conducted for 2 min. at 97°C, 98°C and 99°C, respectively, using graduated concentrations of DNA polymerase.

The amplification procedures used followed approximately the protocol for amplifying nucleic acid sequences outlined by Saiki et al., Science 239:487, 1988. A 1 ml reaction mixture was prepared containing 100 ng of plasmid pLC, 200 pmoles each of primers Lc5 (SEQ ID NO:11) and Lc3 (SEQ ID NO:12),

reaction buffer (20 mM Tris-HCl pH 8.55, 16 mM ammonium sulfate, 2.5 mM MgCl₂ and 150 ug/ml BSA), 200 uM dNTPs, but no enzyme. 100 ul of the reaction mixture was placed into tubes. Aliquots of AmpliTaq and KlenTaq-278 were then added and 20 cycles of PCR were undertaken.

A range of enzyme concentrations was used in order to be able to detect small effects on the effective PCR catalysis activity. The template was 10 ng of pLc (a clone of an R color control cDNA from maize. PNAS 86:7092, Science 247:449). The primers were Lc5 (SEQ ID NO:11) and Lc3 (SEQ ID NO:12).

As a result of this experiment it was seen that 98° C was not detectably detrimental to KlenTaq-278, yet AT was nearly completely inactivated by this temperature.

EXAMPLE 5

Efficient and Accurate PCR Amplification of Long DNA Targets: (Part A)

A preferred embodiment of the above formulation (designated KlenTaq-LA) is provided as follows: Starting with the purified enzymes in storage buffer, mix 1 ul of Pfu DNA polymerase at 2.5 u./ul with 64 ul of KlenTaq-278 at 25 u./ul. Store at -20° C.

Larger amounts of Pfu are detrimental to some PCR amplifications, perform equally for some, and are beneficial for some. For testing of the optimum level of Pfu, several reactions complete with KlenTaq-278 are aliquoted in the amount left to right of 75 ul, 25 ul, 25 ul, and as many additional 25 ul aliquots as desired. Then 3/8 ul of Pfu (equivalent to 0.5 ul per 100 ul -- this is about

the most that one would ever want) is added to the leftmost, 75 ul reaction and mixed. Serial, two-fold dilutions are then made as 25 ul + 25 ul left to right along the row of tubes, adding no Pfu to the last one, as a control of KlenTaq-278 alone. A reaction of 1/2 or 1 ul (per 100 ul) of Pfu alone should also be run.

Reaction buffer is PC2 as above, supplemented with 200 uM of each dNTP and 800 uM of MgCl₂ (total Mg⁺⁺ 3.3 mM), and per 100 ul of reaction volume, 20 pmoles of each primer MBL (SEQ ID NO:7) and MBR (SEQ ID NO:8), and 30 ng of λplac5 intact phage. Per 100 ul of reaction volume, 1 or 1/2 ul of KTLA are effective levels of enzyme. Suitable PCR cycling conditions are two-temperature: 20 seconds at 94° C, 11 minutes at 70° C, for 20 cycles. Alternate cycling conditions include two-temperature PCR with 1 minute at 98° C and 10 minutes at 65° C. 10 to 16 ul are loaded onto an agarose gel for product analysis by staining with ethidium bromide. See Figure 1 for other details and variations. The template was λplac5, which carries a portion of the lac operon region of the E. coli genome. Thirty ng of phage DNA were included in each 100 ul of reaction volume, introduced as intact phage particles. The primers are homologous to wild-type lambda DNA and amplify λ DNA, not the lac DNA. Primer MBL No. 8757 (5' nucleotide matches base pair 27914 of λ DNA) is GCT TAT CTG CTT CTC ATA GAG TCT TGC (SEQ ID NO:7). Primer MBR No. 8835 (5' nucleotide matches bp 34570 of λ DNA) is ATA ACG ATC ATA TAC ATG GTT CTC TCC (SEQ ID NO:8). The size of the amplified product is therefore predicted to be 6657 bp.

As shown in Figure 1A and 1B, each DNA polymerase enzyme (KlenTaq-278 or Pfu) alone gives rise to a faint product band (except for some reactions, when Pfu alone does not work at all), but the combinations all give rise to product bands that are 20 to 50 times more intense than either enzyme can catalyze on its own.

Figure 1C, second lane from the right, shows the surprising result of adding as little as 1/64 ul of Pfu to 1 ul of KlenTaq-278 (a units ratio of 1/640). Not shown are data that as little as 1/200 ul (1/2000 in units) of Pfu contributed a noticeable improvement to the efficiency of this test amplification.

Vent DNA polymerase required 10-fold higher amounts (yet still minority amounts) for similar functionality.

An additional, beneficial, and unexpected attribute to the PCR reactions catalyzed by KlenTaq-LA was a phenomenal, never previously observed intensity and sharpness to the PCR product bands. In part, this increased yield is manifested by a dark area in the middle of the bands as photographed. This darker area in the ethidium fluorescence is believed to be due to UV absorbance by the outside portions of the band, reducing the potential UV-activated fluorescence. The system apparently allowed a much greater yield of product than did the prior art, which tended to create a broad smear of product, and increasing amounts of side product, when amplification was allowed to proceed to this extent.

EXAMPLE 6Efficient and Accurate PCR Amplification of
Long DNA Targets: (Part B)

Efficient amplification of 8.4 kb, 12.5 kb, 15 kb, and 18 kb was demonstrated by the experiment depicted in Figure 2. This experiment extended the demonstrated performance of the a preferred embodiment of the invention, 1/640 KlenTaq-LA, even further. The amplification was highly successful for the size range 8.4 to 15 kb, detectably successful for 18 kb, but not successful for an attempted 19.7 kb.

Eight different PCR reactions were run in this experiment, differing from each other in the template or amount of template or in the primer pair employed, as shown in the legend on Figure 2. Each reaction was divided 3 ways and cycled differently in parts A, B, and C. Between parts A and B, this experiment compared 20 cycles to 30 cycles at 94° denaturation phase. In parts B and C, this experiment compared 94° to 93° for 30 cycles. This experiment utilized 1.3 ul of KlenTaq-LA (at a KlenTaq-278/Pfu ratio of 640) per 100 ul of reaction. This may have been a little too much enzyme, since high enzyme has been associated in previous experiments with the catastrophic synthesis of product which cannot enter the gel, as occurred here for the reaction products in channels 2B and 6C. At the current stage of development of long PCR using the invention, this poor outcome occurs about 10% of the time.

Comparing conditions B and C, it is apparent that a somewhat lower denaturation temperature is desirable. This is consistent with similar experiments comparing

time at 94° C., in which yield of long PCR products was found to be decreased as the denaturation time increased in the order 2, 20, 60, and 180 seconds at 94° C for the denaturation step of each cycle. These data indicate that there was at least one weak link, i.e. least thermostable component, in the reactions which is subject to inactivation at 94°. Since 94° is below the temperature known to damage the DNA polymerase activity and the DNA, it is believed that it is not the thermolabile element. In an alternative embodiment of this aspect of the invention Pfu DNA polymerase is replaced as the minority component with a more thermostable 3'-exonuclease of a DNA polymerase such as, but not limited to, that from the Archaeobacterium strain ES4, which can grow at temperatures up to 114° C [Pledger, R.J. and Baross, J.A., J. Gen. Microbiol. **137** (1991)], which maximum growth temperature exceeds that of the source of the Pfu DNA polymerase (103° C.; Blumentals, I.I. et al. (1990) Annals of the N.Y. Acad. Sci. **589**:301-314.)

In the experiment in Figure 2 the final intensity of the 15 kb band matched in only 20 cycles the yield obtained by Kainze et al. supra in 30 cycles for a band of similar size and from similar λ DNA template amounts. This was a measure of the improved efficiency provided by the invention, and the further result was that the yield catalyzed by the invention in 30 cycles greatly exceeded the yield reported by these authors for 30 cycles. Accurate quantitation has not yet been carried out to measure the efficiency of the two methods, but inspection of Figure 2 compared to the figure published by Kainze et al. shows a yield for the 15 kb fragment that is estimated to be some 100 times higher. This corresponds approximately to a doubled efficiency of PCR extension.

EXAMPLE 7Efficient and Accurate PCR Amplification of
Long DNA Targets: (Part C)Materials and Methods

5 **DNA Polymerases.** DNA polymerases Vent and Deep Vent were supplied by New England Biolabs. Pfu DNA polymerase and its exo⁻ mutant were supplied by Stratagene at 2.5 units/ul. Klentaq-278 is an N-terminal deletion variant of Taq DNA polymerase as described above. Purified Klentaq-278 was as supplied by Ab Peptides, St. Louis, MO, USA at 25-35 units/ul (a protein concentration of about 0.7 ug/ul). One unit of DNA polymerase activity incorporates 10 nmoles of nucleotide in 30 min. at 72° C., utilizing activated (partially degraded) calf thymus DNA as template. Since activated calf thymus DNA is a somewhat undefined substrate and is structurally different from PCR reaction substrate, this assay was routinely eschewed in favor of a PCR-based assay to set the above stock concentration of Klentaq-278: the concentration of Klentaq-278 stock was adjusted so that 0.25 ul effectively (but .12 ul less effectively) catalyzes the amplification of a 2 kb target span from 10 ng of plasmid substrate with cycling conditions including 7 min. of annealing / extension at 65°. The mixture of 15/16 ul Klentaq-278 + 1/16 ul Pfu DNA polymerases is designated KlentaqLA-16.

20 **Agarose gel** electrophoresis employed 0.7% to 1% agarose in 1XGGB (TEA) buffer [40 mM Tris acetate pH 8.3, 20 mM sodium acetate, 0.2 mM EDTA] at 2-3 v/cm, with 3% ficoll instead of glycerol in the loading dye. Figure 5 employed 1% agarose pulsed-field CHEF (11) with a switching time of 4 sec. Standard DNA fragment sizes in every figure are, in kilobases (kb): 23.1, 9.4, 6.6,

4.4, 2.3, 2.0, and 0.56. Figure 5 and 6 also have a full-length λ plac5 standard band, 48645 bp.

All agarose gels were run or stained in ethidium bromide at 0.5 ug/ml and photographed (35 mm ASA 400 black and white film) or videographed (Alpha Innotech or Stratagene Eagle Eye) under UV illumination. While printing the gel photographs, the left halves of Figures 2 and 4 were exposed 50% less than the right halves.

DNA primers are listed in Table 3 and in the Sequence Listing.

Lambda DNA templates. λ vacA, a gift from S. Phadnis, is a λ EMBL4-vectored clone of the cytotoxin gene region of Helicobacter pylori DNA. This DNA was extracted and stored frozen. The other phage template DNAs λ plac5 (12) and λ K138 (13) were added as intact phage particles that had been purified by CsCl equilibrium centrifugation, dialyzed, and diluted in 1X PC2 buffer.

Long and Accurate PCR. PC2 Reaction buffer (10) consisted of 20 mM Tris-HCl pH 8.55 at 25°, 150 ug/ml BSA, 16 mM $(\text{NH}_4)_2\text{SO}_4$, 3.5 mM MgCl_2 , 250 uM each dNTP. For success above 28 kb (at 35 kb), 1.5 ul of 2 M Tris base was added to each reaction, corresponding to pH 9.1 measured for the Tris-HCl component only at 20 mM in water at 25° C. Contact with a pH probe was detrimental to the reactions, so pH was only measured on separate aliquots, and found to be 8.76 in the final reaction at 25° C. Each 100 ul of reaction volume contained 20 pmoles of each primer, and 0.1 to 10 ng of phage DNA template. 0.8 or 1.2 ul of KlentaqLA-16 was appropriate for under 20 kb and over 20 kb, respectively. Reaction volumes per tube were 33-50 ul, under 40 ul of mineral oil in thin-walled (PGC or Stratagene) plastic test tubes.

PCR reactions utilizing the primers at the ends of λ

For reported results of comparison of conditions such as cycling temperatures and times, thermal cycler machines, thick and thin-walled tubes, etc., reactions were made up as 100 ul complete and then split into identical aliquots of 33 ul before subjecting to PCR cycling.

For reported results of comparison of conditions such as cycling temperatures and times, thermal cycler machines, thick and thin-walled tubes, etc., reactions were made up as 100 ul complete and then split into identical aliquots of 33 ul before subjecting to PCR cycling.

Table 3. Primer and template combinations.

Product Size	Left Primer	Right Primer	Template DNA
5.8	MBL101	MS1933	λ K138
6657	MBL	MBR	λ plac5
8386	MBL-1.7	MBR	λ plac5
8.7	MBR001	λ R36	λ K138
12.1	lacZ333	MBR202	λ K138
12.5	MBL 27mer or MBL101 33mer	MBR 27mer or MBR202 33mer	λ vacAI
15560	MSA19 28mer MSA1933 33mer	MBR202	λ plac5
18.0	MBL101	MBR202	λ K138
19.8	L36	MBL002	λ K138
20707	MBL101	λ R36 36mer	λ plac5
19584	λ L36	lacZ333	λ plac5
13971	MBR001 33mer	λ R36	λ plac5
22.0	λ L36	lacZ'533	λ K138
24.6	λ L36	MSA1933	λ K138
22495	λ L36	lacZ536	λ plac5
26194	lacZ533	λ R36	λ plac5
28083	L36	MBL002	λ plac5
34968	L36	MBR202	λ plac5

Legend to Table 3.

Product sizes in integer base pairs are as predicted from the sequence and structure of λ and λ plac5 as documented in Genbank accession no. J02459 and ref. (21). Product sizes with decimal points in kb were determined by comparison with these products and with the λ +HindIII size standards labelled λ H3. The sequence of the primers is given in the Sequence Listing.

Megaprimer consisted of gel-purified 384 bp PCR product DNA homologous to the region between the BamHI site and EcoRI site of the gene coding for the CryV ICP of Bacillus thuringiensis (14), and primer-modified to remove these restriction sites. The PCR reactions in Figure 4 each employed megaprimer (300 ng), primer BtV5 and 20 ng of genomic DNA from Bacillus thuringiensis strain NRD12 (15), and enzyme as indicated in the description above of Figure 3. Cycling

conditions were 30 sec. 95°, 7 min. 60°, for 20 cycles.

HindIII digestion. Unfractionated, total PCR reactions for 28 and 35 kb targets were supplemented with 1/10 volume of 10XNaTMS (1X = 50 mM NaCl, 10 mM Tris-HCl pH 7.7, 10 mM MgCl₂, 10 mM mercaptoethanol) and 2 ul (10 units) of restriction enzyme HindIII, and incubated at 55° C. for 90 min.

Test of exo- Pfu. Each 100 ul of reaction (incubated as 33 ul under 40 ul of oil) contained 2 ng λplac5 DNA as purified phage particles, 20 pmoles each of primers MBL-1.7 and MBR, reaction buffer PC2 and 1 ul of Klentaq-278 (0.7 ug), except for reaction 6, which contained 1 ul Pfu DNA polymerase (2.5 u.) alone. Other details are in the description of Fig. 12. Thermal conditions were 24 cycles of 2 sec. at 94°, 11 min. at 70°.

The discovery leading to the DNA polymerase mixture of the present invention was made during attempts to utilize in PCR a primer with a mismatched A-A base-pair at its 3' end. In fact the primer was itself a PCR product "megaprimer" of 384 bp, and the mismatched A had been added by Klentaq-278 using non-templated terminal transferase activity common to DNA polymerases (16). Neither Klentaq-278 (Figure 3, lane 3) nor Pfu DNA polymerase (Figure 3, lanes 1 & 2 and other levels of enzyme not shown) could catalyze amplification of the 1500 bp target that lay between the PCR-product megaprimer and a 42mer oligonucleotide primer. The combination of the two enzymes, however, was well able to catalyze amplification of the desired target fragment (Fig 3, lane 4). Evidently, the Pfu DNA polymerase removed the presumed 3' A-A mismatch, allowing Klentaq-278 catalysis to proceed efficiently for each step of the PCR. The same result was obtained with Vent DNA polymerase substituted for Pfu (data not shown).

I hypothesized that mismatched 3'-ends are a general cause of inefficient primer extension during PCR of targets larger than a few kb. As a test system I employed a 6.6 kb

lambda DNA target which was amplified detectably but poorly by AmpliTaq, Klentaq-278 or Pfu DNA polymerase in a variety of standard conditions. Per 100 ul reaction volume, 1 ul of Klentaq-278 was combined with various amounts of Pfu DNA polymerase, from 1/2 ul down to as little as 1/200 ul of Pfu. Since the Pfu stock (2.5 units/ul) was at least 10 times less concentrated than the Klentaq-278 stock (25-30 units/ul), the actual ratios tested were 1/20 to 1/2000 in DNA polymerase units. Representative results of these tests are shown in Figure 1B. A high yield of target band was observed for all tested combinations of the two enzymes, yet several levels of each enzyme on its own failed to catalyze more than faintly detectable amplification. The lowest level of Pfu tested, 1/200 ul, exhibited only a slight beneficial effect. The apparent broad optimum ratio of Klentaq-278:Pfu1 was 16 or 64 by volume, which is about 160 or 640 on the basis of DNA polymerase incorporation units. When tested at 6-8 kb (data not shown), other combinations of 3'-exo⁻ and 3'-exo⁺ thermostable DNA polymerases also showed the effect, including Klentaq-278/Vent, Klentaq5 (DeltaTaq, USB) / Pfu, Stoffel Fragment/Pfu, Klentaq-278/Deep Vent (our co-favorite; 48:1 by Volume, 720:1 by unit), and Pfu exo⁻ / Pfu exo⁺. AmpliTaq/Pfu or AmpliTaq/Pwo, at ratio of 25:1, are also very effective, but it is important that the Mg⁺⁺ be held to a level that is close to 0.75mM over the total level of dNTPs [for instance, 400 uM each dNTP, and 2.35 mM MgCl₂.]

A very short heat step is preferred. I next attempted to amplify DNA in the size range 8.4 to 18 kb from lambda transducing phage template. Our early cycling protocol employed a denaturation step of 1 or 2 minutes at 95° or 98° C, but no useful product in excess of 8.4 kb was obtained until the parameters of this heat step were reduced to 2 sec. or 20 sec. at 93° or 94° C. In an experiment with the denaturation step at 94° for 20, 60, or 180 sec, the 8.4 kb product exhibited

decreasing yield with increased length of this heat step (data not shown). Apparently, a component of the reaction is at its margin of thermostability. Figure 2 shows that, using the short 2 sec. denaturation step, target fragment was obtained for some reactions at all sizes in the range 8.4 to 18 kb, with very high product yields up to 15 kb if 30 PCR cycles were employed. Figure 2 also shows some failed reactions which I cannot explain. The failure mode that gives rise to massive ethidium staining in the sample well (30-cycle lane 2) was particularly common, especially at high enzyme levels.

Longer Primers. A change in primer length from 27 to 33 greatly reduced the frequency of failed reactions. Figure 4 demonstrates improved reliability for amplification of 12.5, 15 and 18 kb with the longer 33mer primers, under conditions of otherwise optimally high enzyme levels in which the 27mer primers failed to give rise to desirable target product. This result does not represent an extensive survey of primer length, and it has not yet been repeated with the improvements below. Therefore the optimum primer length for long PCR remains to be determined. Some of the amplifications analyzed in Figure 5 utilized 36mer primers from the very ends of λ . A 2-5 min. preincubation at 68-72° (22) was necessary to release the template DNA from the phage particles and to fill in the sticky ends of lambda to complete the template homology with primers λ L36 and λ R36.

Filtered Tips. For repeated experiments in the same laboratory with the same primer sets, some sort of carry-over product can contaminate the pipetter barrels and stock solutions, and it is now believed that this is the main cause of the failed reactions shown in Figure 4. The nature of the carried-over product has not yet been determined, but it seems to act as a "bad seed" to recruit good PCR product into the intractable material that is at the wells and does not enter the gel in the failed lanes of Figure 4.

This carry-over contamination problem is effectively combatted by two measures: 1) Always use different pipets for assembly (before cycling) and gel analysis (after cycling) procedures. 2) Always use the pipet tips with filters in each one, also known as aerosol resistant tips (ART).

When the above two measures are employed, 27 mer primers and primers as short as 23 base pairs often work well for the long and accurate PCr. When compared directly, 33 mer primers continue to outperform 23 mer primers, but the difference is now slight (less than 3-fold improvement).

Rapid cycling. A change to thin-walled tubes, which have lower heat capacity and conduct heat more efficiently, further improved the reactions. Figure 5 shows a CHEF pulse-field agarose gel analysis of successful amplifications of DNA spans 6-26 kb in size. The target of 28 kb was not amplifiable in the Omnigene thermal cycler (data not shown), but did appear (Figure 6, lane 2) when the RoboCycler was employed.

Several models of thermal cycler have been employed, and although not all have been optimized, some are preferable to others for long PCR. As may be concluded from the advantage of thin-walled tubes noted above, success seems to be positively correlated with a high speed of temperature change made possible by the design of the thermal cycler. The RoboCycler achieves rapid temperature change by moving tubes from block to block, and observations with a thermistor temperature probe indicate that it raises the reactions to 93-95° for only 5 sec. under the denaturation conditions employed (30 sec. in the 99° block), before rapidly (within 30 sec) returning the reaction to 68°.

Higher pH. The current record 35 kb (Figure 6, lane 3) was only amplifiable if the pH was increased. A preliminary scan of higher pH was carried out (data not shown), and this resulted in the appearance of the 35 kb band at pH 8.8 to 9.2, with the optimum at 9.1 as described in Methods (above). Further improvement to a high yield of the 35 kb product was

achieved by lengthening the extension time to 24 min. Other than the higher pH, the long PCR procedure has not yet realized any potential benefits from changes in buffer conditions from those optimized for 8.4 kb. For Targets over 20 kbs extension times exceeding 20 min. are preferred and the extension temperature is preferably below 69° C.

Identity of long PCR products. It can be seen in Figures 2, 4 and 5 that the mobilities of the successful large DNA products agree with those predicted in Table 3 from the known map positions of the primers used.

HindIII restriction enzyme digestion of the unpurified 28 and 35 kb products (Figure 6, lanes 6 and 7) resulted in the expected left arm of lambda (23 kb) and 2.3 kb band from both, and the predictable bands terminated by the right PCR primer: 447 bp (barely visible) from the 28 kb product and 7331 bp from the 35 kb product.

Exonuclease mutant. The available mutant of Pfu DNA polymerase (8) which is defective in the 3'-exonuclease activity was tested. Figure 7 shows that the 3'-exo⁻ mutant of Pfu DNA polymerase fails to promote efficient amplification of a long DNA target. This supports our hypothesis that the 3'-exonuclease activity is important for the efficiency of PCR amplification in this size range.

Fidelity test. Since the biological purpose of 3'-exonuclease is to edit base pair mismatches for high replication fidelity, we tested the fidelity of the PCR product using an assay involving the amplification and molecular cloning of an entire lacZ (β -galactosidase) gene flanked by two selectable markers (10). Heretofore the highest reported fidelity of PCR amplification is that catalyzed by Pfu DNA polymerase (2). Table 4 shows that the fidelity of the product amplified by the 640:1 mixture of KlenTaq-278 and Pfu DNA polymerase at least matches that obtained for Pfu DNA polymerase, alone, when each are used for 16 cycles of PCR. Our designation of the enzyme

mixture as Klentaq-LA (KlenTaq Long and Accurate) reflects this high fidelity performance.

Table 4. Non-silent mutations introduced into the lacZ gene by 16 cycles of PCR (10).

Enzyme	LacZ+ Blue or White	LacZ- Light Blue	% mutant	Effective cycle no. (c)	Errors per 10 ⁵ bp (b)	Fold Improve- ment over full-length Taq
-----	-----	-----	-----	-----	-----	-----
KTLA-64	571	34	5.6	12	1.05	12.7
Pfu	528	37	6.5	8	1.9	6.9
Klentaq5 ^a	442	85	16.1	8	5.1	2.6
Klentaq1	3225	985	26.4	8	9.0	1.5
Amplitaq	525	301	36.4	8	13.4	1.0

(a) Klentaq⁵ is the N-terminal deletion of Taq DNA polymerase described in ref. 10. (b) Equation 1 of reference 10 was rearranged to be as follows to solve for errors per bp: $X = -(\ln(2F^{(1/m-1)} - 1))/1000$, where X is the errors per bp incorporated, 1000 is the effective target size in the lacZ gene (10), F is the fraction of blue colonies, and m is the effective cycle number. (c) As in ref. 10, the effective cycle number was estimated at less than the machine cycles to reflect the actual efficiency of the reaction, yet higher than the minimum calculated from the fold-amplification. Strand loss due to incomplete synthesis of product strands is a probable cause of lower than ideal amplification efficiency. Therefore successful (not lost) product molecules are judged to have undergone more than the calculated minimum number of replications. KTLA-64 (Klentaq-278:Pfu::64:1 by volume) was assigned a higher effective cycle number since its reactions started with 10 times less DNA (1.5 ng vs. 15 ng plasmid pWB305) to result in comparable levels of product.

DISCUSSION

FOI 20018550

5 The previous length limitation for PCR amplification is postulated to have been caused by low efficiency of extension at the sites of incorporation of mismatched base pairs. Although it would have seemed that the cure for these mismatches would be to employ enzymes with 3'-(editing)-exonucleases, I believe that when Pfu and Vent DNA polymerase are used to catalyze our amplifications on their own, their failure is due to degradation of the PCR primers by their 3'-exonucleases, especially during the required long synthesis times and at optimally high DNA polymerase levels. Evidently, low levels of 3'-exonuclease are sufficient and optimal for removal of the mismatches to allow the KlenTaq-278 and amplification to proceed. It has been demonstrated that the optimally low level of 3'-exonuclease can be set effectively, conveniently, and flexibly by mixing and dilution.

20 Preferably the ratio of exo-/exo+ enzyme is high. If equal levels of the two types of enzymes are used (or where the E2 component is in excess), and in many embodiments tested, where the ratio of exo-/exo+ is 4 or less, the effectiveness of the long PCR, even under optimal cycling conditions discussed below, is non-existent or much reduced.

25 It is preferred, and for certain applications, important that the length and temperature of the heat denaturation step of the PCR be kept to a minimum. Further, the improvement obtained by increasing the pH slightly may correspond to a decrease in template depurination. If so, further improvements may result if depurination can be reduced, or if a majority DNA polymerase component can be found which is able to bypass depurination sites.

30 The short denaturation time found to be optimal, preferably less than 20 sec., and most preferably, 5 sec. or less in the reaction itself at 95°, is surprisingly effective for the amplification of 35 kb, whereas it might have been expected that longer PCR targets would need longer denaturation

time to become completely denatured. If complete denaturation is required for PCR, and if longer DNA requires more time to unwind at 95°, the required unwinding time may eventually become significantly more than 5 seconds. This could limit the size of amplifiable product because of the increased depurination caused by longer denaturation times.

These amplifications were successful with several different target sequences, with several primer combinations, and with product sizes up to nearly twice the maximum size of inserts cloned into λ . Whole viruses and plasmids up to 35 kb in length should now be amplifiable with this system. Should this method prove applicable to DNA of higher complexity than λ , it could prove a boon to genomic mapping and sequencing applications, since in vitro amplification is convenient and avoids the DNA rearrangement and gene toxicity pitfalls of in vivo cloning.

In view of the above, it will be seen that the several objects of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and products without departing from the scope of the invention, it is intended that all matter contained in the above description shall be interpreted as illustrative and not in a limiting sense.

References

1. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988) Science **239**, 487-491.
- 5 2. Lundberg, K.S., Shoemaker, D.D., Adams, M.W.W., Short, J.M. Sorge, J.A., and Mathur, E.J. (1991) Gene **108**, 1-6.
3. Lawyer, F.C., Stoffel, S., Saiki, R.K., Chang, S-Y., Landre, P.A., Abramson, R.D., and Gelfand, D.H. (1993) PCR Methods and Applications **2**, 275-287.
4. Jeffreys, A.J., Wilson, V., Neumann, R., and Keyte, J. (1988) Nucleic Acids Res. **16**, 10953-10971.
5. Krishnan, B.R., Kersulyte, D., Brikun, I., Berg, C.M. & Berg, D.E. (1991) Nucleic Acids Res. **19**, 6177-6182.
6. Maga, E.A., & Richardson, T. (1991) BioTechniques **11**: 185-186.
7. Ohler, L.D. & Rose, E. A. (1992) PCR Methods and Applications **2**, 51-59.
8. Rychlik, W., Spencer, W.J., and Rhoads, R.E. (1990) Nucleic Acids Res. **18**: 6409.
- 20 9. Kainz, P. Schmiedlechner, A., & Strack, H.B. (1992) Analytical Biochem. **202**, 46-49.
10. Barnes, W.M. (1992) Gene **112**, 29-35.
11. Chu, G.D., Vollrath, D. and Davis, R.W. (1986) Science **234**, 1582-.
- 25 12. Ippen, K., Shapiro, J.A., and Beckwith, J.R. (1971) J.Bact. **108**, 5-9.
13. Kohara, Y. (1990) pp 29-42 in The Bacterial Chromosome, eds Drlica, K. & Riley, M., ASM Washington D.C.
- 30 14. Tailor, R., Tippet, J., Gibb, G., Pells, S., Pike, D., Jordan, L., Ely, S. (1992) Mol. Microbiol. **6**, 1211-1217.
15. Dubois, N.R. Reardon, R.C., Kolodny-Hirsh, D.M. J.Econ.Entomol. **81**, 1672 (1988)],18.
16. Clark, J.M. (1988) Nucleic Acids Res. **16**, 9677-.
- 35 17. Brewer, A.C., Marsh, P.J and Patient, R.K. (1990) Nucleic Acids Res. **18**, 5574.

0981002 201701

18. Lindahl, T. (1993) Nature 362: 709-715.
19. Lindahl, T. & Nyberg, B. (1972) Biochemistry 11: 3611-3618.
20. Sigma Chemical Co. Technical Bulletin No. 106B.
21. Shpakovski, G.V., Akhrem, A.A., and Berlin, Y.A. (1988)
Nucleic Acids Res 16, 10199 (1988).

5

T02F01" 200F8660

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 09981002-1010
- (i) APPLICANT: Barnes Ph.D., Wayne M
- (ii) TITLE OF INVENTION: DNA polymerases with enhanced length and efficiency of primer extension
- (iii) NUMBER OF SEQUENCES: 29
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Senniger, Powers, Leavitt & Roedel
(B) STREET: One Metropolitan Square
(C) CITY: St. Louis
(D) STATE: Missouri
(E) COUNTRY: USA
(F) ZIP: 63102
- Sub B 7
(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Blosser, G. Harley
(B) REGISTRATION NUMBER: 33,650
(C) REFERENCE/DOCKET NUMBER: WNB4912
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (314) 231-5400
(B) TELEFAX: (314) 231-4342
(C) TELEX: 6502697583 MCI

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: N-terminal
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *Thermus aquaticus*
 - (B) STRAIN: YT1
- (vii) IMMEDIATE SOURCE:
 - (A) LIBRARY: synthetic
 - (B) CLONE: KT1
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 6..35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAGCC ATG GGC CTC CTC CAC GAG TTC GGC CTT CTG G
36
Met Gly Leu Leu His Glu Phe Gly Leu Leu
1 5 10

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Gly Leu Leu His Glu Phe Gly Leu Leu
1 5 10

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: YES
- (v) FRAGMENT TYPE: C-terminal
- (vi) ORIGINAL SOURCE:
 (A) ORGANISM: *Thermus aquaticus*
 (B) STRAIN: YT1
- (vii) IMMEDIATE SOURCE:
 (A) LIBRARY: synthetic
 (B) CLONE: Klentaq32
- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: complement (8..34)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
- GCGAAGCTTA CTACTCCTTG GCGGAGAGCC AGTCC

FEQVOT"200F8660

5

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asp Trp Leu Ser Ala Lys Glu * *

1 5

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6714 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: circular

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Expression vector
- (vii) IMMEDIATE SOURCE:
(B) CLONE: pWB254b
- (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1665

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATG GGC CTC CTC CAC GAG TTC GGC CTT CTG GAA AGC CCC AAG GCC CTG 48
Met Gly Leu Leu His Glu Phe Gly Leu Leu Glu Ser Pro Lys Ala Leu
5 10 15

GAG GAG GCC CCC TGG CCC CCG CCG GAA GGG GCC TTC GTG GGC TTT GTG 96
Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly Ala Phe Val Gly Phe Val
20 25 30

CTT TCC CGC AAG GAG CCC ATG TGG GCC GAT CTT CTG GCC CTG GCC GCC 144
Leu Ser Arg Lys Glu Pro Met Trp Ala Asp Leu Leu Ala Leu Ala Ala
35 40 45

GCC AGG GGG GGC CGG GTC CAC CGG GCC CCC GAG CCT TAT AAA GCC CTC 192
Ala Arg Gly Gly Arg Val His Arg Ala Pro Glu Pro Tyr Lys Ala Leu
50 55 60

AGG GAC CTG AAG GAG GCG CGG GGG CTT CTC GCC AAA GAC CTG AGC GTT 240
Arg Asp Leu Lys Glu Ala Arg Gly Leu Leu Ala Lys Asp Leu Ser Val
65 70 75 80

CTG GCC CTG AGG GAA GGC CTT GGC CTC CCG CCC GGC GAC GAC CCC ATG 288
Leu Ala Leu Arg Glu Gly Leu Gly Leu Pro Pro Gly Asp Asp Pro Met
85 90 95

CTC CTC GCC TAC CTC CTG GAC CCT TCC AAC ACC ACC CCC GAG GGG GTG 336
Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn Thr Thr Pro Glu Gly Val
100 105 110

GCC CGG CGC TAC GGC GGG GAG TGG ACG GAG GAG GCG GGG GAG CGG GCC 384
Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu Glu Ala Gly Glu Arg Ala
115 120 125

GCC Ala	CTT Leu	TCC Ser	GAG Glu	AGG Arg	CTC Leu	TTC Phe	GCC Ala	AAC Asn	CTG Leu	TGG Trp	GGG Gly	AGG Arg	CTT Leu	GAG Glu	GGG Gly	432
130						135					140					
GAG Glu	GAG Glu	AGG Arg	CTC Leu	CTT Leu	TGG Trp	CTT Leu	TAC Tyr	CGG Arg	GAG Glu	GTG Val	GAG Glu	AGG Arg	CCC Pro	CTT Leu	TCC Ser	480
145					150					155					160	
GCT Ala	GTC Val	CTG Leu	GCC Ala	CAC His	ATG Met	GAG Glu	GCC Ala	ACG Thr	GGG Gly	GTG Val	CGC Arg	CTG Leu	GAC Asp	GTG Val	GCC Ala	528
				165					170					175		
TAT Tyr	CTC Leu	AGG Arg	GCC Ala	TTG Leu	TCC Ser	CTG Leu	GAG Glu	GTG Val	GCC Ala	GAG Glu	GAG Glu	ATC Ile	GCC Ala	CGC Arg	CTC Leu	576
			180					185					190			
GAG Glu	GCC Ala	GAG Glu	GTC Val	TTC Phe	CGC Arg	CTG Leu	GCC Ala	GGC Gly	CAC His	CCC Pro	TTC Phe	AAC Asn	CTC Leu	AAC Asn	TCC Ser	624
		195					200					205				
CGG Arg	GAC Asp	CAG Gln	CTG Leu	GAA Glu	AGG Arg	GTC Val	CTC Leu	TTT Phe	GAC Asp	GAG Glu	CTA Leu	GGG Gly	CTT Leu	CCC Pro	GCC Ala	672
	210					215					220					
ATC Ile	GGC Gly	AAG Lys	ACG Thr	GAG Glu	AAG Lys	ACC Thr	GGC Gly	AAG Lys	CGC Arg	TCC Ser	ACC Thr	AGC Ser	GCC Ala	GCC Ala	GTC Val	720
225					230					235					240	
CTG Leu	GAG Glu	GCC Ala	CTC Leu	CGC Arg	GAG Glu	GCC Ala	CAC His	CCC Pro	ATC Ile	GTG Val	GAG Glu	AAG Lys	ATC Ile	CTG Leu	CAG Gln	768
				245					250					255		
TAC Tyr	CGG Arg	GAG Glu	CTC Leu	ACC Thr	AAG Lys	CTG Leu	AAG Lys	AGC Ser	ACC Thr	TAC Tyr	ATT Ile	GAC Asp	CCC Pro	TTG Leu	CCG Pro	816
			260					265					270			
GAC Asp	CTC Leu	ATC Ile	CAC His	CCC Pro	AGG Arg	ACG Thr	GGC Gly	CGC Arg	CTC Leu	CAC His	ACC Thr	CGC Arg	TTC Phe	AAC Asn	CAG Gln	864
		275					280					285				
ACG Thr	GCC Ala	ACG Thr	GCC Ala	ACG Thr	GGC Gly	AGG Arg	CTA Leu	AGT Ser	AGC Ser	TCC Ser	GAT Asp	CCC Pro	AAC Asn	CTC Leu	CAG Gln	912
	290					295					300					
AAC Asn	ATC Ile	CCC Pro	GTC Val	CGC Arg	ACC Thr	CCG Pro	CTT Leu	GGG Gly	CAG Gln	AGG Arg	ATC Ile	CGC Arg	CGG Arg	GCC Ala	TTC Phe	960
305					310					315					320	
ATC Ile	GCC Ala	GAG Glu	GAG Glu	GGG Gly	TGG Trp	CTA Leu	TTG Leu	GTG Val	GCC Ala	CTG Leu	GAC Asp	TAT Tyr	AGC Ser	CAG Gln	ATA Ile	1008
				325					330					335		

GAG Glu	CTC Leu	AGG Arg	GTG Val 340	CTG Leu	GCC Ala	CAC His	CTC Leu	TCC Ser 345	GGC Gly	GAC Asp	GAG Glu	AAC Asn	CTG Leu 350	ATC Ile	CGG Arg	1056
GTC Val	TTC Phe	CAG Gln 355	GAG Glu	GGG Gly	CGG Arg	GAC Asp	ATC Ile 360	CAC His	ACG Thr	GAG Glu	ACC Thr	GCC Ala 365	AGC Ser	TGG Trp	ATG Met	1104
TTC Phe 370	GGC Gly	GTC Val	CCC Pro	CGG Arg	GAG Glu	GCC Ala 375	GTG Val	GAC Asp	CCC Pro	CTG Leu	ATG Met 380	CGC Arg	CGG Arg	GCG Ala	GCC Ala	1152
AAG Lys 385	ACC Thr	ATC Ile	AAC Asn	TTC Phe	GGG Gly 390	GTC Val	CTC Leu	TAC Tyr	GGC Gly	ATG Met 395	TCG Ser	GCC Ala	CAC His	CGC Arg	CTC Leu 400	1200
TCC Ser 400	CAG Gln	GAG Glu	CTA Leu	GCC Ala 405	ATC Ile	CCT Pro	TAC Tyr	GAG Glu	GAG Glu 410	GCC Ala	CAG Gln	GCC Ala	TTC Phe	ATT Ile 415	GAG Glu	1248
CGC Arg 420	TAC Tyr	TTT Phe	CAG Gln 420	AGC Ser	TTC Phe	CCC Pro	AAG Lys 425	GTG Val 425	CGG Arg	GCC Ala	TGG Trp	ATT Ile 430	GAG Glu 430	AAG Lys	ACC Thr	1296
CTG Leu 430	GAG Glu	GAG Glu 435	GGC Gly	AGG Arg	AGG Arg	CGG Arg	GGG Gly 440	TAC Tyr	GTG Val	GAG Glu	ACC Thr	CTC Leu 445	TTC Phe	GGC Gly	CGC Arg	1344
CGC Arg 450	CGC Arg	TAC Tyr	GTG Val	CCA Pro	GAC Asp	CTA Leu 455	GAG Glu	GCC Ala	CGG Arg	GTG Val	AAG Lys 460	AGC Ser	GTG Val	CGG Arg	GAG Glu	1392
GCG Ala 465	GCC Ala	GAG Glu	CGC Arg	ATG Met	GCC Ala 470	TTC Phe	AAC Asn	ATG Met	CCC Pro	GTC Val 475	CAG Gln	GGC Gly	ACC Thr	GCC Ala 480	GCC Ala	1440
GAC Asp	CTC Leu	ATG Met	AAG Lys 485	CTG Leu	GCT Ala	ATG Met	GTG Val	AAG Lys 490	CTC Leu	TTC Phe	CCC Pro	AGG Arg	CTG Leu 495	GAG Glu 495	GAA Glu	1488
ATG Met	GGG Gly	GCC Ala	AGG Arg 500	ATG Met	CTC Leu	CTT Leu	CAG Gln	GTC Val 505	CAC His	GAC Asp	GAG Glu	CTG Leu	GTC Val 510	CTC Leu	GAG Glu	1536
GCC Ala	CCA Pro	AAA Lys 515	GAG Glu	AGG Arg	GCG Ala	GAG Glu	GCC Ala 520	GTG Val	GCC Ala	CGG Arg	CTG Leu	GCC Ala 525	AAG Lys	GAG Glu	GTC Val	1584
ATG Met 530	GAG Glu	GGG Gly	GTG Val	TAT Tyr	CCC Pro	CTG Leu 535	GCC Ala	GTG Val	CCC Pro	CTG Leu	GAG Glu 540	GTG Val	GAG Glu	GTG Val	GGG Gly	1632

ATA GGG GAG GAC TGG CTC TCC GCC AAG GAG TAGTAAGCTT ATCGATGATA	1682
Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu	
545 550 555	
AGCTGTCAAA CATGAGAATT AGCCCGCCTA ATGAGCGGGC TTTTTTTTAA TTCTTGAAGA	1742
CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT AATGGTTTCT	1802
TAGCGTCAAA GCAACCATAG TACGCGCCCT GTAGCGGCGC ATTAAGCGCG CCGGGTGTGG	1862
TGGTTACGCG CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT	1922
TCTTCCCTTC CTTTCTCGCC ACGTTCGCCG GCTTTCCTCCG TCAAGCTCTA AATCGGGGGC	1982
TCCGCTTTAGG GTTCCGATTT AGTGCTTTAC GGCACCTCGA CCCCAGAAAAA CTTGATTTGG	2042
GTGATGGTTC ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG	2102
AGTCCACGTT CTTTAATAGT GGACTCTTGT TCCAAACTTG AACAACTC AACCCTATCT	2162
CGGGCTATTC TTTTGATTTA TAAGGGATTT TGCCGATTTT GGCCTATTGG TTAAAAAATG	2222
AGGTGATTTA ACAAAAATTT AACGCGAATT TTAACAAAAT ATTAACGTTT ACAATTTTCAG	2282
GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT	2342
CAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA	2402
GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT	2462
GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT	2522
TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT	2582
TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG	2642
TATTATCCCG TGTTGACGCC GGGCAAGAGC AACTCGGTCT CCGCATACAC TATTCTCAGA	2702
ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA	2762
GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA	2822
CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA	2882
CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA	2942
CCACGATGCC TGCAGCAATG GCAACAACGT TGCGCAAAT ATTAAGTGGC GAACTACTTA	3002
CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC	3062
TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC	3122

GTGGGTCTCG	CGGTATCATT	GCAGCACTGG	GGCCAGATGG	TAAGCCCTCC	CGTATCGTAG	3182
TTATCTACAC	GACGGGGAGT	CAGGCAACTA	TGGATGAACG	AAATAGACAG	ATCGCTGAGA	3242
TAGGTGCCTC	ACTGATTAAG	CATTGGTAAC	TGTCAGACCA	AGTTTACTCA	TATATACTTT	3302
AGATTGATTT	AAAAC TTCAT	TTTTAATTTA	AAAGGATCTA	GGTGAAGATC	CTTTTTTGATA	3362
ATCTCATGAC	CAAAATCCCT	TAACGTGAGT	TTTCGTTCCA	CTGAGCGTCA	GACCCCGTAG	3422
AAAAGATCAA	AGGATCTTCT	TGAGATCCTT	TTTTTCTGCG	CGTAATCTGC	TGCTTGCAAA	3482
CAAAAAAACC	ACCGCTACCA	GCGGTGGTTT	GTTTGCCGGA	TCAAGAGCTA	CCA ACTCTTT	3542
TTCCGAAGGT	AACTGGCTTC	AGCAGAGCGC	AGATACCAAA	TACTGTCCTT	CTAGTGTAGC	3602
CGTAGTTAGG	CCACCACTTC	AAGAACTCTG	TAGCACCGCC	TACATACCTC	GCTCTGCTAA	3662
TCCGTTTACC	AGTGGCTGCT	GCCAGTGGCG	ATAAGTCGTG	TCTTACCGGG	TTGGACTCAA	3722
GACGATAGTT	ACCGGATAAG	GCGCAGCGGT	CGGGCTGAAC	GGGGGGTTCG	TGCACACAGC	3782
CCAGCTTGGA	GCGAACGACC	TACACCGAAC	TGAGATACCT	ACAGCGTGAG	CTATGAGAAA	3842
GCGGCACGCT	TCCCGAAGGG	AGAAAGGCGG	ACAGGTATCC	GGTAAGCGGC	AGGGTCGGAA	3902
CAGGAGAGCG	CACGAGGGAG	CTTCCAGGGG	GAAACGCCTG	GTATCTTTAT	AGTCCTGTCTG	3962
GGTTTCGCCA	CCTCTGACTT	GAGCGTCGAT	TTTTGTGATG	CTCGTCAGGG	GGGCGGAGCC	4022
TATGGAAAAA	CGCCAGCAAC	GCGGCCTTTT	TACGGTTCCT	GGCCTTTTGC	TGGCCTTTTG	4082
CTCACATGTT	CTTTCCTGCG	TTATCCCCTG	ATTCTGTGGA	TAACCGTATT	ACCGCCTTTG	4142
AGTGAGCTGA	TACCGCTCGC	CGCAGCCGAA	CGACCGAGCG	CAGCGAGTCA	GTGAGCGAGG	4202
AAGCGGAAGA	GCGCCTGATG	CGGTATTTTC	TCCTTACGCA	TCTGTGCGGT	ATTTACACACC	4262
GCATATGGTG	CACTCTCAGT	ACAATCTGCT	CTGATGCCGC	ATAGTTAAGC	CAGTATACAC	4322
TCCGCTATCG	CTACGTGACT	GGGTCATGGC	TGCGCCCCGA	CACCCGCCAA	CACCCGCTGA	4382
CGCGCCCTGA	CGGGCTTGTC	TGCTCCCGGC	ATCCGCTTAC	AGACAAGCTG	TGACCGTCTC	4442
CGGGAGCTGC	ATGTGTCAGA	GGTTTTACAC	GTCATCACCG	AAACGCGCGA	GGCAGAACGC	4502
CATCAAAAAT	AATTCGCGTC	TGGCCTTCCT	GTAGCCAGCT	TTCATCAACA	TTAAATGTGA	4562
GCGAGTAACA	ACCCGTCGGA	TTCTCCGTGG	GAACAAACGG	CGGATTGACC	GTAATGGGAT	4622
AGGTTACGTT	GGTGTAGATG	GGCGCATCGT	AACCGTG CAT	CTGCCAGTTT	GAGGGGACGA	4682

CGACAGTATC	GGCCTCAGGA	AGATCGCACT	CCAGCCAGCT	TTCCGGCACC	GCTTCTGGTG	4742
CCGGAAACCA	GGCAAAGCGC	CATTGCGCAT	TCAGGCTGCG	CAACTGTTGG	GAAGGGCGAT	4802
CGGTGCGGGC	CTCTTCGCTA	TTACGCCAGC	TGGCGAAAGG	GGGATGTGCT	GCAAGGCGAT	4862
TAAGTTGGGT	AACGCCAGGG	TTTTCCCAGT	CACGACGTTG	TAAAACGACG	GCCAGTGAAT	4922
CCGTAATCAT	GGTCATAGCT	GTTTCCTGTG	TGAAATTGTT	ATCCGCTCAC	AATTCCACAC	4982
AACATACGAG	CCGGAAGCAT	AAAGTGTAAG	GCCTGGGGTG	CCTAATGAGT	GAGCTAACTC	5042
ACATTAATTG	CGTTGCGCTC	ACTGCCCCGCT	TTCCAGTCGG	GAAACCTGTC	GTGCCAGCTG	5102
CATTAATGAA	TCGGCCAACG	CGCGGGGAGA	GGCGGTTTGC	GTATTGGGCG	CCAGGGTGGT	5162
TTTCTTTTC	ACCAGTGAGA	CGGGCAACAG	CTGATTGCCC	TTCACCGCCT	GGCCCTGAGA	5222
GAGTTGCAGC	AAGCGGTCCA	CGCTGGTTTG	CCCCAGCAGG	CGAAAATCCT	GTTTGATGGT	5282
GGTTGACGGC	GGGATATAAC	ATGAGCTGTC	TTCGGTATCG	TCGTATCCCA	CTACCGAGAT	5342
ATCCGCACCA	ACGCGCAGCC	CGGACTCGGT	AATGGCGCGC	ATTGCGCCCA	GCGCCATCTG	5402
ATCGTTGGCA	ACCAGCATCG	CAGTGGGAAC	GATGCCCTCA	TTCAGCATTT	GCATGGTTTG	5462
TTGAAAACCG	GACATGGCAC	TCCAGTCGCC	TTCCCGTTCC	GCTATCGGCT	GAATTTGATT	5522
GCGAGTGAGA	TATTTATGCC	AGCCAGCCAG	ACGCAGACGC	GCCGAGACAG	AACTTAATGG	5582
GCCCCGTAAC	AGCGCGATTT	GCTGGTGACC	CAATGCGACC	AGATGCTCCA	CGCCCAGTCG	5642
CGTACCGTCT	TCATGGGAGA	AAATAATACT	GTTGATGGGT	GTCTGGTCAG	AGACATCAAG	5702
AAATAACGCC	GGAACATTAG	TGCAGGCAGC	TTCCACAGCA	ATGGCATCCT	GGTCATCCAG	5762
CGGATAGTTA	ATGATCAGCC	CACTGACGCG	TTGCGCGAGA	AGATTGTGCA	CCGCCGCTTT	5822
ACAGGCTTCG	ACGCCGCTTC	GTTCTACCAT	CGACACCACC	ACGCTGGCAC	CCAGTTGATC	5882
GGCGCGAGAT	TTAATCGCCG	CGACAATTTG	CGACGGCGCG	TGCAGGGCCA	GACTGGAGGT	5942
GGCAACGCCA	ATCAGCAACG	ACTGTTTGCC	CGCCAGTTGT	TGTGCCACGC	GGTTGGGAAT	6002
GTAATTCAGC	TCCGCCATCG	CCGCTTCCAC	TTTTTCCCGC	GTTTTTCGCAG	AAACGTGGCT	6062
GGCCTGGTTC	ACCACGCGGG	AAACGGTCTG	ATAAGAGACA	CCGGCATACT	CTGCGACATC	6122
GTATAACGTT	ACTGGTTTCA	CATTCAACCAC	CCTGAATTGA	CTCTCTTCCG	GGCGCTATCA	6182
TGCCATACCG	CGAAAGGTTT	TGCGCCATTC	GATGGTGTCC	CAGTGAATCC	GTAATCATGG	6242

TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAT TATACGAGCC 6302
 GGAAGCATAA AGTGTAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC ATTAATTGCG 6362
 TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA TTAATGAATC 6422
 GGAGCTTACT CCCCATCCCC CTGTTGACAA TTAATCATCG GCTCGTATAA TGTGTGGAAT 6482
 TGTGAGCGGA TAACAATTTT ACACAGGAAA CAGGATCGAT CCAGCTTACT CCCCATCCCC 6542
 CTGTTGACAA TTAATCATCG GCTCGTATAA TGTGTGGAAT TGTGAGCGGA TAACAATTTT 6602
 ACACAGGAAA CAGGATCTGG GCCCTTCGAA ATTAATACGA CTCACTATAG GGAGACCACA 6662
 ACGGTTTCCC TCTAGAAATA ATTTTGTTTA ACTTTAAGAA GGAGATATAT CC 6714

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 554 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Gly Leu Leu His Glu Phe Gly Leu Leu Glu Ser Pro Lys Ala Leu
 1 5 10 15
 Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly Ala Phe Val Gly Phe Val
 20 25 30
 Leu Ser Arg Lys Glu Pro Met Trp Ala Asp Leu Leu Ala Leu Ala Ala
 35 40 45
 Ala Arg Gly Gly Arg Val His Arg Ala Pro Glu Pro Tyr Lys Ala Leu
 50 55 60
 Arg Asp Leu Lys Glu Ala Arg Gly Leu Leu Ala Lys Asp Leu Ser Val
 65 70 75 80
 Leu Ala Leu Arg Glu Gly Leu Gly Leu Pro Pro Gly Asp Asp Pro Met
 85 90 95
 Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn Thr Thr Pro Glu Gly Val
 100 105 110
 Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu Glu Ala Gly Glu Arg Ala
 115 120 125

Ala Leu Ser Glu Arg Leu Phe Ala Asn Leu Trp Gly Arg Leu Glu Gly
 130 135 140
 Glu Glu Arg Leu Leu Trp Leu Tyr Arg Glu Val Glu Arg Pro Leu Ser
 145 150 155 160
 Ala Val Leu Ala His Met Glu Ala Thr Gly Val Arg Leu Asp Val Ala
 165 170 175
 Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala Glu Glu Ile Ala Arg Leu
 180 185 190
 Glu Ala Glu Val Phe Arg Leu Ala Gly His Pro Phe Asn Leu Asn Ser
 195 200 205
 Arg Asp Gln Leu Glu Arg Val Leu Phe Asp Glu Leu Gly Leu Pro Ala
 210 215 220
 Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg Ser Thr Ser Ala Ala Val
 225 230 235 240
 Leu Glu Ala Leu Arg Glu Ala His Pro Ile Val Glu Lys Ile Leu Gln
 245 250 255
 Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr Tyr Ile Asp Pro Leu Pro
 260 265 270
 Asp Leu Ile His Pro Arg Thr Gly Arg Leu His Thr Arg Phe Asn Gln
 275 280 285
 Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser Ser Asp Pro Asn Leu Gln
 290 295 300
 Asn Ile Pro Val Arg Thr Pro Leu Gly Gln Arg Ile Arg Arg Ala Phe
 305 310 315 320
 Ile Ala Glu Glu Gly Trp Leu Leu Val Ala Leu Asp Tyr Ser Gln Ile
 325 330 335
 Glu Leu Arg Val Leu Ala His Leu Ser Gly Asp Glu Asn Leu Ile Arg
 340 345 350
 Val Phe Gln Glu Gly Arg Asp Ile His Thr Glu Thr Ala Ser Trp Met
 355 360 365
 Phe Gly Val Pro Arg Glu Ala Val Asp Pro Leu Met Arg Arg Ala Ala
 370 375 380
 Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly Met Ser Ala His Arg Leu
 385 390 395 400

Ser Gln Glu Leu Ala Ile Pro Tyr Glu Glu Ala Gln Ala Phe Ile Glu
 405 410 415
 Arg Tyr Phe Gln Ser Phe Pro Lys Val Arg Ala Trp Ile Glu Lys Thr
 420 425 430
 Leu Glu Glu Gly Arg Arg Arg Gly Tyr Val Glu Thr Leu Phe Gly Arg
 435 440 445
 Arg Arg Tyr Val Pro Asp Leu Glu Ala Arg Val Lys Ser Val Arg Glu
 450 455 460
 Ala Ala Glu Arg Met Ala Phe Asn Met Pro Val Gln Gly Thr Ala Ala
 465 470 475 480
 Asp Leu Met Lys Leu Ala Met Val Lys Leu Phe Pro Arg Leu Glu Glu
 485 490 495
 Met Gly Ala Arg Met Leu Leu Gln Val His Asp Glu Leu Val Leu Glu
 500 505 510
 Ala Pro Lys Glu Arg Ala Glu Ala Val Ala Arg Leu Ala Lys Glu Val
 515 520 525
 Met Glu Gly Val Tyr Pro Leu Ala Val Pro Leu Glu Val Glu Val Gly
 530 535 540
 Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
 545 550

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: bacteriophage lambda
- (B) STRAIN: PaPa

(vii) IMMEDIATE SOURCE:

- (B) CLONE: MBL

(viii) POSITION IN GENOME:

- (B) MAP POSITION: 27940

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCTTATCTGC TTTCATAGA GTCTTGC

27

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: bacteriophage lambda
- (B) STRAIN: PaPa

(vii) IMMEDIATE SOURCE:

- (B) CLONE: MBR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATAACGATCA TATACATGGT TCTCTCC

27

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: bacteriophage lambda

(vii) IMMEDIATE SOURCE:

- (B) CLONE: MBL-1.7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTTTGCTGGG TCAGGTTGTT CTTTAGG

27

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(v) FRAGMENT TYPE: C-terminal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: E.coli
- (B) STRAIN: K12

(vii) IMMEDIATE SOURCE:

- (B) CLONE: MSA19

(viii) POSITION IN GENOME:

- (B) MAP POSITION: lacZ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGAAGCTTAT TTTTGACACC AGACCAAC

28

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Zea maize

(vii) IMMEDIATE SOURCE:

- (B) CLONE: Lc5

(viii) POSITION IN GENOME:

- (B) MAP POSITION: 5' end of color control gene Lc

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTGATGGATC CTTGAGCTTC CCGAGTTCAG CAGGCGG

37

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iv) ANTI-SENSE: YES

(v) FRAGMENT TYPE: C-terminal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Zea maize

(vii) IMMEDIATE SOURCE:
(B) CLONE: Lc3

(viii) POSITION IN GENOME:
(B) MAP POSITION: 3' end of color control gene Lc

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGTCTCGAGC GAAGCTTCCC TATAGCTTTG CGAAGAG

37

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Thermus aquaticus
(B) STRAIN: YT1

(vii) IMMEDIATE SOURCE:
(B) CLONE: KT2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAGCCATGGC CAACCTGTGG GGGAGGCTTG AGGGGGA

37

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *Thermus aquaticus*
 - (B) STRAIN: YT1

- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: Genbank Accession no. J04639

- (viii) POSITION IN GENOME:
 - (A) CHROMOSOME/SEGMENT: DNA polymerase gene
 - (B) MAP POSITION: 950

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGTTTGGCAG CCTCCTCCAC GAGTTCGGCC TTCTGG

36

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *Thermus aquaticus*
 - (B) STRAIN: YT1

- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: Genbank Accession No. J04639

- (viii) POSITION IN GENOME:
 - (A) CHROMOSOME/SEGMENT: DNA polymerase gene
 - (B) MAP POSITION: 2595

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGACTGGCTC TCCGCCAAGG AGTGATACCA CC

32

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Thermus flavus*

(vii) IMMEDIATE SOURCE:
(B) CLONE: Genbank Accession No. X66105

(viii) POSITION IN GENOME:
(A) CHROMOSOME/SEGMENT: DNA polymerase
(B) MAP POSITION: 1378

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AGTATGGAAG CCTCCTCCAC GAGTTCGGCC TCCTGG

36

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *Thermus flavus*

(vii) IMMEDIATE SOURCE:
(B) CLONE: Genbank Accession No. X66105

(viii) POSITION IN GENOME:
(A) CHROMOSOME/SEGMENT: DNA polymerase gene
(B) MAP POSITION: 3023

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GGACTGGCTC TCCGCCAAGG AGTAGGGGGG TCCTG

35

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Bacillus thuringiensis*
- (B) STRAIN: CryV
- (C) INDIVIDUAL ISOLATE: NRD12

(vii) IMMEDIATE SOURCE:

- (B) CLONE: BtV3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCGAAGCTTC TCGAGTTACG CTCAATATGG AGTTGCTTC

39

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Bacillus thuringiensis*
- (B) STRAIN: NRD12

(vii) IMMEDIATE SOURCE:

- (B) CLONE: BtV5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCGAGATCTC CATGGATCCA AAGAATCAAG ATAAGCATCA AAG

43

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: bacteriophage lambda
 - (B) STRAIN: PaPa

- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: L36

- (viii) POSITION IN GENOME:
 - (B) MAP POSITION: left end

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGGCGGCGAC CTCGCGGGTT TTCGCTATTT ATGAAA

36

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(v) FRAGMENT TYPE: N-terminal

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: E.coli
 - (B) STRAIN: K12

- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: lacZ'533

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGACGGCCAG TGAATCCGTA ATCATGGTCA TAG

33

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (v) FRAGMENT TYPE: C-terminal
- (vi) ORIGINAL SOURCE:
 (A) ORGANISM: E.coli
 (B) STRAIN: K12
- (vii) IMMEDIATE SOURCE:
 (B) CLONE: lacZ333
- (viii) POSITION IN GENOME:
 (B) MAP POSITION: lacZ
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ACCAGCCATC GCCATCTGCT GCACGCGGAA GAA

33

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iv) ANTI-SENSE: NO
- (v) FRAGMENT TYPE: N-terminal
- (vi) ORIGINAL SOURCE:
 (A) ORGANISM: E.coli
 (B) STRAIN: K12
- (vii) IMMEDIATE SOURCE:
 (B) CLONE: lacZ536
- (viii) POSITION IN GENOME:
 (B) MAP POSITION: lacZ
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CTATGACCAT GATTACGGAT TCACTGGCCG TCGTTT

36

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: bacteriophage lambda
- (B) STRAIN: PaPa

(vii) IMMEDIATE SOURCE:

- (B) CLONE: MBL002

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCAAGACTCT ATGAGAAGCA GATAAGCGAT AAG

33

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: bacteriophage lambda
- (B) STRAIN: PaPa

(vii) IMMEDIATE SOURCE:

- (B) CLONE: MBL101

(viii) POSITION IN GENOME:

- (B) MAP POSITION: 27840

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ATCATTATTT GATTTC AATT TTGTCCCACT CCC

33

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: bacteriophage lambda
 - (B) STRAIN: PaPa

- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: MBR001

- (viii) POSITION IN GENOME:
 - (B) MAP POSITION: 34576

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGAGAGAACC ATGTATATGA TCGTTATCTG GGT

33

- (2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: bacteriophage lambda
 - (B) STRAIN: PaPa

- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: MBR202

- (viii) POSITION IN GENOME:
 - (B) MAP POSITION: 34793

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCGCACAAAA CCATAGATTG CTCTTCTGTA AGG

33

- (2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(v) FRAGMENT TYPE: C-terminal

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: E.coli
- (B) STRAIN: K12

(vii) IMMEDIATE SOURCE:

- (B) CLONE: MSA1933

(viii) POSITION IN GENOME:

- (B) MAP POSITION: lacZ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CCCGTTATT ATTATTTTGTG ACACCAGACC AAC

33

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: bacteriophage lambda
- (B) STRAIN: PaPa

(vii) IMMEDIATE SOURCE:

- (B) CLONE: R36

(viii) POSITION IN GENOME:

- (B) MAP POSITION: right end

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AGGTCGCCGC CCCGTAACCT GTCGGATCAC CGGAAA

36